DOCUMENT RESUME

ED 336 423 TM 017 236

TITLE	Oversight Hearings on Educational Assessment. Hearings before the Subcommittee on Elementary, Secondary, and Vocational Education. Committee on Education and Labor, House of Representatives, One Hundred Second Congress, First Session (March 13-14, 1991).
INSTITUTION	Congress of the U.S., Washington, D.C. House Committee on Education and Labor.
PUB DATE	91
NOTE	368p.; Serial No. 102-8.
PUB TYPE	Legal/Legislative/Regulatory Materials (090)
EDRS PRICE	MF01/PC15 Plus Postage.
DESCRIPTORS	*Educational Assessment; Educational Change; Educational Improvement; Educational Trends; Elementary Secondary Education; Futures (of Society); Government Role; Hearings; *National Programs; Teacher Evaluation; *Testing Programs
IDENTIFIERS	Congress 102nd; *National Assessment of Educational Progress

ABSTRACT

The focus of this 2-day hearing was on the National Assessment of Educational Progress (NAEP). Witnesses testified about the importance of the NAEP as a national assessment tool and the changes that the Subcommittee on Elementary, Secondary, and Vocational Education (SESVE) might consider for the future. On March 13, the following persons testified: H. J. Walberg; D. M. Koretz; R. L. Linn; and E. D. Roeber. On March 14, the following persons addressed the SESVE: R. Romer; W. E. Brock; G. R. Anrig; T. E. Sewell; S. Cooperman; K. V. Hertz; M. Neill; E. W. Kelley; and G. M. Ambach. Debate over the hearings ranged beyond the NAEP to consider the concept of a national test and a national testing program for teachers. The remarks of all contributors are followed by the prepared statements, letters, and attachments that they submitted to the SESVE. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED336423

OVERSIGHT HEARINGS ON EDUCATIONAL ASSESSMENT

HEARINGS

BEFORE THE

SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION

OF THE

COMMITTEE ON EDUCATION AND LABOR
HOUSE OF REPRESENTATIVES

ONE HUNDRED SECOND CONGRESS

FIRST SESSION

HEARINGS HELD IN WASHINGTON, DC, MARCH 13 AND 14, 1991

Serial No. 102-8

Printed for the use of the Committee on Education and Labor

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

BEST COPY AVAILABLE

TM017236

U.S. GOVERNMENT PRINTING OFFICE
41-090
WASHINGTON : 1991

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, DC 20402

COMMITTEE ON EDUCATION AND LABOR

WILLIAM D. FORD, Michigan, *Chairman*

JOSEPH M. GAYDOS, Pennsylvania
WILLIAM (BILL) CLAY, Missouri
GEORGE MILLER, California
AUSTIN J. MURPHY, Pennsylvania
DALE E. KILDEE, Michigan
PAT WILLIAMS, Montana
MATTHEW G. MARTINEZ, California
MAJOR R. OWENS, New York
CHARLES A. HAYES, Illinois
CARL C. PERKINS, Kentucky
THOMAS C. SAWYER, Ohio
DONALD M. PAYNE, New Jersey
NITA M. LOWEY, New York
JOLENE UNSOELD, Washington
CRAIG A. WASHINGTON, Texas
JOSÉ E. SERRANO, New York
PATSY T. MINK, Hawaii
ROBERT E. ANDREWS, New Jersey
WILLIAM J. JEFFERSON, Louisiana
JOHN F. REED, Rhode Island
TIM ROEMER, Indiana
PETER J. VISCLOSKY, Indiana
RON DE LUGO, Virgin Islands
JAIME B. FUSTER, Puerto Rico

WILLIAM F. GOODLING, Pennsylvania
E. THOMAS COLEMAN, Missouri
THOMAS E. PETRI, Wisconsin
MARGE ROUKEMA, New Jersey
STEVE GUNDERSON, Wisconsin
RICHARD K. ARMEY, Texas
HARRIS W. FAWELL, Illinois
PAUL B. HENRY, Michigan
CASS BALLENGER, North Carolina
SUSAN MOLINARI, New York
BILL BARRETT, Nebraska
JOHN A. BOEHNER, Ohio
SCOTT L. KLUG, Wisconsin

SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION

DALE E. KILDEE, Michigan, *Chairman*

GEORGE MILLER, California
PAT WILLIAMS, Montana
MATTHEW G. MARTINEZ, California
CARL C. PERKINS, Kentucky
CHARLES A. HAYES, Illinois
THOMAS C. SAWYER, Ohio
MAJOR R. OWENS, New York
NITA M. LOWEY, New York
JOLENE UNSOELD, Washington
WILLIAM J. JEFFERSON, Louisiana
JOHN F. REED, Rhode Island
TIM ROEMER, Indiana
CRAIG A. WASHINGTON, Texas
PATSY T. MINK, Hawaii
JAIME B. FUSTER, Puerto Rico
WILLIAM D. FORD, Michigan
(Ex Officio)

WILLIAM F. GOODLING, Pennsylvania
SCOTT L. KLUG, Wisconsin
THOMAS E. PETRI, Wisconsin
MARGE ROUKEMA, New Jersey
STEVE GUNDERSON, Wisconsin
PAUL B. HENRY, Michigan
SUSAN MOLINARI, New York
JOHN A. BOEHNER, Ohio

CONTENTS

Page

Hearings held in Washington, DC:

- March 13, 1991 1
- March 14, 1991 135

Statement of:

- Cooperman, Saul, President, Educate America, Morristown, NJ; Karl V. Hertz, Superintendent, Mequon-Thiensville School District, Mequon, WI; Monty Neill, Associate Director, FairTest, Cambridge, MA; E.W. Kelley, Professor of American Government, Cornell University, Ithaca, NY; and Gordon M. Ambach, Executive Director, Council of Chief State School Officers, Washington, DC 206
- Romer, Hon. Roy, Chair, National Education Goals Panel, Denver, CO; Senator William E. Brock, Chairman, Secretary's Commission on the Achievement of Necessary Skills, Washington, DC; Gregory R. Anrig, President, Educational Testing Service, Princeton, NJ; and Trevor Sewell, Acting Dean, College of Education, Temple University, Philadelphia, PA 139
- Walberg, Herbert J., Member, National Assessment Governing Board, Chicago, IL; Dr. Daniel M. Koretz, Senior Social Scientist, RAND Corp., Washington, DC; Dr. Robert L. Linn, Professor, University of Colorado, Boulder, CO; and Dr. Edward D. Roeber, Supervisor, Assessment and Accreditation, Michigan Department of Education, Lansing, MI 3

Prepared statements, letters, supplemental materials, et cetera:

- Akers, John F., Chairman and CEO, IBM Corp., prepared statement on behalf of The Business Roundtable 125
- Ambach, Gordon M., Executive Director, Council of Chief State School Officers, Washington, DC, prepared statement of 272
- Anrig, Gregory R., President, Educational Testing Service, Princeton, NJ, prepared statement of 161
- Association of American Publishers, Inc., prepared statement of 324
- Brock, Senator William E., Chairman, Secretary's Commission on the Achievement of Necessary Skills, Washington, DC, prepared statement of 152
- Cooperman, Saul, President, Educate America, Morristown, NJ, prepared statement of 209
- Hertz, Karl V., Superintendent, Mequon-Thiensville School District, Mequon, WI, prepared statement of 215
- Houghton Mifflin Company, prepared statement of 355
- Kelley, E.W., Professor of American Government, Cornell University, Ithaca, NY, prepared statement of 244
- Koretz, Dr. Daniel M., Senior Social Scientist, RAND Corporation, Washington, DC:
  - Letter dated March 15, 1991, from Hon. Dale E. Kildee, enclosing additional questions for the record, and responses to same 113
  - Prepared statement of 42
- Linn, Dr. Robert L., Professor, University of Colorado, Boulder, CO:
  - Letter dated March 15, 1991, from Hon. Dale E. Kildee, enclosing additional questions for the record, and responses to same 119
  - Prepared statement of 56
- Middle Cities Education Association, prepared statement of 352
- National Alliance of Business, prepared statement of 308
- National Education Association, prepared statement of 298
- Neill, Monty, Associate Director, FairTest, Cambridge, MA
  - Letter dated March 18, 1991, to Hon. Dale Kildee, enclosing additional material for the record 127

Page
Prepared statements, letters, supplemental materials, et cetera—Continued
  Neill, Monty, Associate Director, FairTest, Cambridge, MA—Continued
    Prepared statement of 231
  Owens, Hon. Major R., a Representative in Congress from the State of New York, prepared statement of 137
  Roeber, Dr. Edward D., Supervisor, Assessment and Accreditation, Michigan Department of Education, Lansing, MI:
    Letter dated March 15, 1991, from Hon. Dale E. Kildee, enclosing additional questions for the record, and responses to same 92
    Prepared statement of 65
  Romer, Hon. Roy, Chair, National Education Goals Panel, Denver, CO, prepared statement of 143
  Sewell, Trevor, Acting Dean, College of Education, Temple University, Philadelphia, PA, prepared statement of 178
  United States Catholic Conference, prepared statement of 306
  Walberg, Herbert J., Member, National Assessment Governing Board, Chicago, IL:
    Letter dated March 15, 1991, from Hon. Dale E. Kildee, enclosing additional questions for the record, and responses to same 97
    Prepared statement of 8

OVERSIGHT HEARING ON THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP)

WEDNESDAY, MARCH 13, 1991

House of Representatives
Subcommittee on Elementary, Secondary,
and Vocational Education,
Committee on Education and Labor,
Washington, DC.

The subcommittee met, pursuant to call, at 9 a.m., Room 2257, Rayburn House Office Building, Hon. Dale E. Kildee [Chairman] presiding.

Members present: Representatives Kildee, Martinez, Sawyer, Goodling, and Gunderson.

Chairman Kildee. The Subcommittee on Elementary, Secondary, and Vocational Education will come to order.

The subcommittee meets today for the first of several hearings on educational testing. Today's hearing focuses on the National Assessment of Educational Progress.

As the only federally mandated assessment of student achievement, NAEP has become the focus of much attention in the national testing debate. Today's witnesses bring a high level of expertise and will advise us regarding NAEP's importance as a national assessment tool and what changes the committee might consider in the future.

Today's testimony will also provide a foundation for tomorrow's hearing when the subcommittee will examine the pros and cons of national testing generally.

Certainly the question of what, if any, NAEP's role should or will be is an integral part of this discussion. I had hoped that the administration would provide a witness this morning to explain how NAEP is administered, and while the Department of Education declined at this time the subcommittee's invitation to testify, it is my understanding that Secretary Designate Lamar Alexander has a great deal of interest—we know that—in educational testing.

I am informed that he would like to testify on NAEP himself at some later point, and I look forward to that testimony.

My opening statement will be brief this morning. I turn to my friend and the ranking Republican member of the committee and also the ranking Republican member of the full committee, Bill Goodling of Pennsylvania.

Mr. GOODLING. Thank you, Mr. Chairman. I would like to commend you for calling the hearings today and tomorrow on the important topic of educational testing and assessment.

With national goals, national tests, and national curriculum, and last year the big kick was national certification of teachers—all these great things we are going to do to improve education—being discussed seriously across the country, it is important that this committee become actively involved in this area.

I would like to put into the record, and read it into the record, a statement of my Secretary of Education back in the State of Pennsylvania in a letter that he wrote, and I won't say to whom he wrote the letter.

"If the speed of the invitations to give testimony on national testing indicates interest, as you suggest, it also indicates a pell mell rush to jump on a band wagon. I do not have time to develop my response beyond the following bullets but wish to have my concerns on record.

The principles of restructured education—flexibility, changed student-teacher roles, classroom based management, et cetera—cannot be accomplished for the long term cast against a national testing program. I don't care how sensitively it is constructed and administered; it will ultimately drive our education train, not just guide and inform it.

There is in my mind little or no relationship between the improvement of student learning and a national testing program. There is a relationship between student performance on national standards and a national testing program, but they are not the same thing. National testing is the wrong way to go if we believe that schools are places for learning, not sorting kids into successes or failures. No matter what is intended by such a program, I believe one inevitable result will be to place academic value signs on individual children. The attached article states many of my thoughts on the issue."

I just want to read one or two of those.

"Yes, everyone is concerned about the quality of our schools, and yes, everyone wants more accountability from our schools, but there is already a multitude of national standardized tests. How will yet another national exam help improve elementary and secondary education? The Ford Foundation reported recently that mandatory testing consumes some 20 million school days and the equivalent of $700 million to $900 million in direct and indirect expenditures annually, an enormous cost and use of classroom time that could be spent on skill development.

Here are seven reasons why national tests won't improve our schools. One, they do nothing to help students learn. Two, they are poor indicators of student performance. Three, they provide no in-depth assessment of the curriculum in a particular school district. Four, they encourage teaching to the test. Five, they categorize and label students. Six, they are racially, culturally, and socially biased. Seven, they measure trivial information easily tested by multiple choice."

I would tell you, Mr. Chairman, this Secretary has a pretty good reason for being rather concerned. His predecessor in my State had the grand idea that we would test all the public school children, and then we would rate the schools according to the test results and come up with this magnificent idea that a school district in western Pennsylvania is number one in the State of Pennsylvania.

Of course, I supervised student teachers in that school district many years ago. If it isn't number one, they sure aren't getting their money's worth because, number one, they have probably double per pupil expenditure than many other school districts in the State of Pennsylvania.

This isn't the fault of a State formula. It is just that they add to what they get from the State and are willing to pay, and can easily pay much, much more.

Then, secondly, every parent is a college graduate, many of whom have Ph.D.'s, and thirdly, those children by the time they open their eyes or even before, the demand that the parents place on excellence from their children as far as education is concerned and from the schools—I mean, it is just automatic so that is why you would get his real concern about whatever it is we may be doing as far as any kind of national test is concerned.

Sorry I took that much time, but I thought it was important. I may not be here tomorrow to get all those comments in.

Chairman **Kildee**. No, I appreciate very much your opening statement, Mr. Goodling. It is very helpful to the committee and our own personal conversations have always been very helpful to me.

Our panel this morning will consist of Dr. Herbert J. Walberg, board member, National Assessment Governing Board, Chicago, Illinois; Dr. Daniel M. Koretz, Senior Social Scientist, RAND Corporation, Washington, DC; Dr. Robert L. Linn, Professor, University of Colorado, Boulder, Colorado; and from my own State of Michigan Dr. Edward D. Roeber, Supervisor, Assessment and Accreditation, Michigan Department of Education, Lansing, Michigan.

If they would step forward, and if they would testify in the order in which I called their names or any order they may have devised among themselves.

Dr. Walberg, though, will lead off.

STATEMENTS OF HERBERT J. WALBERG, MEMBER, NATIONAL ASSESSMENT GOVERNING BOARD, CHICAGO, ILLINOIS; DR. DANIEL M. KORETZ, SENIOR SOCIAL SCIENTIST, RAND CORPORATION, WASHINGTON, DC; DR. ROBERT L. LINN, PROFESSOR, UNIVERSITY OF COLORADO, BOULDER, COLORADO; AND DR. EDWARD D. ROEBER, SUPERVISOR, ASSESSMENT AND ACCREDITATION, MICHIGAN DEPARTMENT OF EDUCATION, LANSING, MICHIGAN

Dr. **Walberg**. Mr. Chairman, let me express my appreciation for the opportunity to testify before this committee. I would like to request that my full written testimony be made part of the record.

Chairman **Kildee**. Yes, without objection you may summarize in any fashion you wish.

Dr. **Walberg**. Thank you. I will try to be brief, but I have been asked by staff to give an overview of the National Assessment, and I will try to do that as quickly as I can.

As you indicated, my name is Herb Walberg. I am a professor of education at the University of Illinois at Chicago, and one of my areas of specialty is testing and measurement.

I think, as you know, the National Assessment is our only nationally representative assessment of what American students know and can do. It has a 20-year record of providing what we hope is useful information to educators, the public, parents, and others that are concerned about how our nation is doing in terms of the standard school subjects such as reading, mathematics, science, and even adult literacy as well.

My remarks this morning will give something of an overview, but I would like to stick pretty much to board policy. If any questions come up, I would be glad to give my own personal view, but I would like to represent what our board has been doing and what we have been considering.

The main purpose of the National Assessment is to improve our nation's schools by making objective information available about student performance in selected learning areas, available to policymakers at the national, regional, State, and local levels.

We report on students at roughly ages 9, 13, and 17 which correspond to grades 4, 8, and 12. I think it is important to recognize about the National Assessment that not only has it been providing this information, I think, with considerable utility to educators, policymakers, legislators, and others, but it does this in a way that minimizes testing time.

Although people think sometimes that the National Assessment involves mass testing, there are design features in the National Assessment especially having to do with sampling that enables us to restrict this to very small numbers of students, and we have never proposed to do mass tests of a whole State or the whole Nation. Other groups are now proposing that, but we have never done that.

Like national public opinion surveys such as the Gallop Poll in which, say, about 1500 people can provide a reasonably good estimate of what Americans are thinking, we by taking small, representative samples can make a very good estimate of what our students are doing from a national standpoint.

We also make considerable effort to minimize the student testing time for an individual student. We do not take more than an hour of any individual's test time so we have a considerable amount of effort to do that.

One of our main roles, unlike almost any other testing program in the United States, is implied in one of the terms that we use, "progress." We are very interested in measuring national trends over time. That enables us to find out if students are improving or not improving, what kinds of students, and in the past we had regional assessments of different parts of the country.

As you know now, by Congressional legislation we are testing specific States that want to participate in the program on a strictly voluntary basis. Roughly 40 States participate on that voluntary basis on the first round, and in the second proposed round another 40 have also voluntarily chosen to participate in the new assessment.

The role of the National Assessment Governing Board is to be the policymaking board. The board sets policy with respect to such things as the frequency of the assessment, the subjects that will be tested, and certain design features.

There is a three, or tripartite—it is a governing structure in which the board sets policy. The National Center for Educational Statistics as a part of the United States Department of Education gives oversight and monitoring of that policy, and then there are contract grants and other contractual arrangements. The vendors carry out the actual work. Under the 1988 reauthorization Congress created this tripartite structure.

One of the things that I think is particularly important and may be of interest to you as you consider national testing, particularly with respect to the National Assessment, is the independence of the board. That has been a matter of some concern to us.

The board, in its Congressional foundation, emphasized that it was an independent board. It consists of two governors, two State legislators, three teachers, a secondary school principle, an elementary school principle, a local and State superintendent, two curriculum specialists, two testing specialists, one non-public school administrator, a representative of business and industry, and three representatives of the general public so it is very much of a general board, and the idea of this board, we think, in the Congressional legislation was to keep it separate from the Department of Education and various kinds of special interest groups that might have a special agenda with respect to national curriculum.

Two of the chief things that we have been working on since the inception were mandated by the Congressional legislation. One of the most important is what we call a consensus process.

In order to come up with an examination that is fair to students in different parts of the country, we try to gather together groups that would have a special expertise: teachers, other educators, people in business and other fields, that formulate a prescription or goals about what should be taught, and this is a very painstaking process. It takes a lot of information and analysis to look at all the new curricula that are being tried out in American schools and come up with a consensus on what ought to be in that test. So it is a very painstaking process, and we have been working very carefully to make a faithful representation of what is actually being taught in American schools.

The other big effort that we have been making since our inception is to set achievement levels, and again this was by—we were asked to do this in the Congressional legislation.

In the past 18 or 19 years, the National Assessment simply reported on how students were doing such as the averages of students in various parts of the country, but for the first time we have been asked to set standards and so we have gotten groups similar to this consensus type groups, educators and people from other walks of life in the United States, to say not only how the students are doing but how well they are doing, so for the first time we will have national standards.

We have tried this out for the first time so we will be able to say what percentage of our students are actually proficient—proficient meaning, in general, that they are able to go on to the subject matter of the next grade level—so we will be able to say how many of our students are able to do that, taking it somewhat beyond that, and I could go into more detail later if that is of interest to you.

We have conducted particularly this first round, and especially with respect to State comparisons very much on a trial basis. It is called trial in the legislation, and we are doing this as an experiment. We are trying to get through the voluntary 40 States that are participating an estimate of how each State is doing, and then we will be comparing the States, not only with respect to the averages of the different States but with respect to whether they meet these standards.

We hope that this will be useful to people in local school districts, to States, and in the Nation, particularly not only to know how well they are doing but what characteristics of their programs may be accounting for the differences between the States.

To give an example, if we had 10 States that had, let's say, effective homework policies, for example, or particular types of teacher training programs, we would like to know how well those States have done in comparison with other States. We would not be able to conclude with certainty that certain policies caused those things, but people would provide one source of information on whether some of their policies appear to be effective or not.

And as you know, because of the extraordinary interest in educational reform and since 1983 with the publication of A Nation At Risk, many States on a State basis are trying some extraordinary policies of one kind or another, and they would like to see how they would rank, so that is the interest in these types of comparisons.

We would like to take two further initiatives, and perhaps might be considered in further legislation. This is fully described in the documents that I have submitted to the committee, but we would like to authorize further State assessments.

We have done the first—we have plans, conducted the first assessment and the second has been authorized, but we would like authorization, since there has been a keen interest in this, if—and I should mention, too, that our work is being carefully evaluated and assessed not only but by various other groups, the National Academy of Education, the National Center for Educational Statistics, and there has been a lot of interest in the scholarly community and the public policy community in how well these things are doing—we would like authorization if these steps that we are taking to make State comparisons are successful, we would like to continue them further and continue to improve them.

A second point that we would like to bring before you is the possibility of conducting the National Assessment on an annual basis rather than a biennial basis. We think because of the extraordinary amount of interest, certainly in Congress but also by the educational summit of President Bush and the governors that have stated national goals, that we need to have this testing more frequent so we can trace specific progress from year to year.

Also we think, on the board, that we need to clarify the board's independence. There is explicit language in the authorizing legislation that the board is to, and I quote, "formulate the policy guidelines for the national assessment" and to, quote, "exercise its functions, powers, and duties independent of the Secretary and other offices and officers of the Department of Education."

Yet the legislation also says that under the name legislation the Commissioner of Educational Statistics carries out with the advice of the National Assessment Governing Board, and we think that there is a discrepancy there as to whether we are truly a governing board as our name implies, or whether we are merely advisory.

So we would like for the committee to consider those factors.

Well, I have perhaps taken a bit more time than I should. I think some of these issues are quite fundamental. I would be pleased to answer any questions that you have, either now or in writing later or whenever the chairman is pleased.

[The prepared statement of Herbert J. Walberg follows:]

THE NATION'S REPORT CARD naep
National Assessment Governing Board
National Assessment of Educational Progress

TESTIMONY OF

HERBERT J. WALBERG

BEFORE THE

SUBCOMMITTEE ON

ELEMENTARY, SECONDARY AND VOCATIONAL EDUCATION

COMMITTEE ON EDUCATION AND LABOR

MARCH 13, 1991

1100 L Street, N.W.
Suite 7322
Washington, D.C. 20005-4013
(202) 357-6938

Mr. Chairman, let me express my appreciation for the opportunity to testify before the Subcommittee on Elementary, Secondary and Vocational Education about the National Assessment of Educational Progress (NAEP).

My name is Herbert J. Walberg. I have been a member of the National Assessment Governing Board (the Board) since its inception in October of 1988 and am chairman of its Design and Analysis Committee. I am also professor of education at the University of Illinois; one of my areas of specialty is testing and measurement.

NAEP, in its twenty-two year history has developed a well deserved reputation for integrity and innovation. Now known as "The Nation's Report Card," NAEP is our only nationally representative assessment of what America's students know and can do. Also, as part of a pilot project authorized by the Congress in 1988, NAEP will report in June 1991 for the first time State-representative results from the 1990 mathematics assessment.

I am honored to provide testimony on this esteemed federal program. However, I would want to emphasize that my remarks about possible changes to the NAEP legislation will be limited to policies adopted by the Board. As you know, the NAEP legislation specifies the responsibilities and functions of the Board. The law also stipulates that "...the Board...shall be independent of the Secretary [of Education] and the other offices and officers of the Department of Education." Thus, my remarks will not necessarily represent positions of the Department of Education.

This morning I will provide an overview of NAEP and the role of the Board, discuss the utility of State-representative data, and describe Board policies that require legislative action.

OVERVIEW OF NAEP

The purpose of NAEP is "...to improve our Nation's schools by making objective information about student performance in selected learning areas available to policymakers at the national, regional, state and local levels."

Under current law, the National Assessment is conducted every two years. The subjects covered and their prescribed cycles are: reading and mathematics, at least once every two years; writing and science, at least once every four years; and history/geography at least once every six years. The law also permits the Board to select other subject areas for assessment, including assessments of adult literacy.

The National Assessment reports results on students at ages nine, thirteen and seventeen and in grades four, eight and twelve. However, it is important to point out that minimizing test burden is an overriding NAEP policy. NAEP does not test all students at these ages and grades, it tests only a small sample. For example, of the more than nine million students in the three grades covered by NAEP, approximately 51,000 participated in the newly developed 1990 reading assessment and approximately 2,000 students per State participated in the 1990 eighth grade mathematics State trial assessment.

Another way NAEP limits test burden is by minimizing student testing time. Testing time under NAEP is about one hour per tested student. This is accomplished by a procedure known as matrix sampling. In matrix sampling, equivalent subsamples of students receive different portions of the total pool of test questions. The total student sample is large enough to estimate precise results for each question. Since NAEP is intended to report aggregate student results at the national and regional levels and, under the trial State assessment, at the State level, this is an efficient way to assess student performance.

One final aspect of NAEP should be highlighted--its capacity both to report trends over time and to incorporate new developments in curriculum and in assessment technology. To do this, NAEP conducts parallel assessments, one for trend data and one that is "cross-sectional," to reflect current thinking and practices. For trends, NAEP uses procedures and items from past assessments; this maintains what is now a twenty-year trend line in subjects such as reading and mathematics. The cross-sectional assessment reflects current research and best practice in each subject area identified through a national consensus process that is administered by the Board. I will be talking more about this consensus process later in my testimony.

ROLE OF THE NATIONAL ASSESSMENT GOVERNING BOARD

In 1988 the Congress reauthorized NAEP, making improvements, providing for the State trial assessments, requiring evaluations, and redesigning the governance structure to include an independent Board to formulate policy for NAEP.

Prior to 1988, NAEP was conducted under a grant from the Department of Education. The pre-1988 NAEP authority required the grantee to establish an Assessment Policy Committee (APC) to advise on policy matters. This arrangement was intended to insulate NAEP policy development and the conduct of the assessment from the Department of Education. But it also had the effect of impeding the Department's ability to assure accountability in the administration of the NAEP program.

Under the 1988 NAEP reauthorization, the Congress created a tripartite structure for NAEP: an independent governing board to formulate policy for the assessment; administration of the NAEP program within the National Center for Education Statistics; and conduct of NAEP through contracts, grants or cooperative agreements. The result is a system of checks and balances that permits the Department of Education to remain at "arms length" from testing policy but allows appropriate accountability for program administration.

The legislative history of the current NAEP legislation indicates that the independence of the Board was an important consideration; both the Senate and Conference reports discuss provisions that have this purpose. At least three provisions of the law appear to be directed at the independence of the Board; specifically, the provision already mentioned regarding the Board's independence of the Secretary; the requirement that the Board "exercise its independent judgment free from inappropriate influences and special interests"; and the requirement that appointments to the Board are made by the Secretary of Education only from candidates nominated by the Board.

In addition, the composition of the Board as prescribed in the law seems to be founded on the idea of an independent Board. The Board has twenty-three voting members: two governors and two state legislators (both on a bipartisan basis); three teachers; an elementary and a secondary school principal; a local and a State school board member; two chief state school officers; one school superintendent; two curriculum specialists; two testing experts; one nonpublic school administrator; a representative of business or industry; and three representatives of the general public. The only federal member, the Assistant Secretary for Educational Research and Improvement, serves in a non-voting capacity.

It is difficult to imagine a group of individuals who, by the nature of their private roles, would be more grounded in the principle of State and local primacy in education matters and less inclined toward federal interventionism.

We believe that a 1987 report about NAEP was considered by the Congress when it took up the 1988 NAEP reauthorization. The report of the Study Group headed by then Governor Lamar Alexander and H. Thomas James--entitled "The Nation's Report Card"--recommended a tripartite governance structure similar to that in current law. With respect to the body that would be responsible for NAEP policy, the Study Group recommended that:

> The governance and policy direction of the national assessment should be furnished by a broadly representative [Board] that provides wisdom, stability and continuity; that is charged with meshing the assessment needs of states and localities with that of the nation; that is accountable to the public--and to the federal government--for stewardship of this

> important activity; but that is itself buffered from manipulation by any individual, level of government, or special interest within the field of education.

With this as background, I would like to discuss some of the specific responsibilities of the Board as set out in law.

Formulating Policy

The general responsibility of the Board is to formulate the policy guidelines for the National Assessment. This includes formulating policy for: the methodology of the assessment; analyzing data and reporting results; selecting non-mandated subjects for assessment; and improving the form and use of the National Assessment.

In addition, the Board has responsibility for two activities that are relatively large in scope and important to the conduct of NAEP. The first of these, forming the foundation for each assessment, is the work of developing consensus on the content for each assessment. The second, identifying achievement levels for each grade for each assessment, in fact establishes benchmarks for interpreting the results.

Consensus Process on Test Content

Developing consensus on the content of each assessment involves much work. With a decentralized public education system of 16,000 school districts and fifty-odd state education agencies (including the District of Columbia and the outlying areas), the need to include non-public schools as well, and the importance of incorporating the perspectives of researchers, practitioners, and subject area specialists, arriving at a consensus in each subject area on what should be included in NAEP is an immense and complicated undertaking. It involves literally hundreds of people, is completed over a period of twelve to eighteen months, and costs approximately $450,000 per subject area.

The result of this consensus activity is a document that, for each grade, describes in very specific detail what the test will include. This document is used by the NAEP contractor to develop the test questions that will be used for the assessment in that subject area.

Achievement Levels

Likewise, setting achievement levels for each assessment is a large and complicated undertaking, and at the present time, the subject of intense scrutiny. Setting achievement levels, or defining what students should know and be able to do in each grade and subject tested under NAEP, was required for the first time under the 1988 NAEP reauthorization. Unlike consensus on test content, setting achievement levels has never previously been done.

Realizing that it was in uncharted territory, the Board has attempted to proceed prudently. The Board decided that its first effort would be a pilot limited to one subject--in this case mathematics as part of the 1990 assessment. The Board has obtained the advice of the foremost experts in the field of education standard-setting to design the methodology for setting achievement levels. The actual work of setting the achievement levels involves the participation of both educators and non-educators. In addition, the Board has conducted three public hearings to receive comment at various points in the process.

Currently, the Board is conducting a validation/replication study of the initial work on achievement levels for the 1990 mathematics assessment. And concurrent with these activities is an external evaluation of the process. Overall, setting achievement levels will involve hundreds of individuals over a period of about twelve months at a cost of approximately $350,000.

The result of this first achievement level-setting activity will be a document that describes what students should know and be able to do in mathematics at the fourth, eighth and twelfth grades that will be used, on a pilot basis, for interpreting results on the 1990 mathematics assessment. Called for in the congressional legislation, we intend to build upon this experience so that this ambitious experimental effort can be improved in subsequent assessments. As a consequence of our experience thus far, we are already developing ways to improve the process.

NAEP REAUTHORIZATION ISSUES

Over the last fifteen months, the Board has adopted a number of policies that would require legislative action. These policies and the rationale for each are described in detail at attachment A. I would like to highlight four of those policies.

The first two issues address a common theme: use of NAEP for reporting results at the State and below the State (i.e. district and school level). The Board views these two issues differently, and I will discuss each in some detail.

The Board believes that State-representative and school- and district-representative results, properly used, can be very helpful in informing policy and in improving instruction.

For example, in addition to average overall scores for the 1990 eighth grade mathematics State trial assessment, there will be subscores by content area (e.g. numbers and operations, measurement, geometry, data analysis, and algebra). In addition, background questions in areas such as teacher certification, inservice training, instructional practices, and parental involvement are also part of the assessment.

Thus, a State superintendent of instruction might compare State results on the subscales with those of a similar or neighboring State, and at the same time look at teacher certification and inservice training data. This kind of comparison can help identify strengths as well as weaknesses, identify States that may have useful models, but most importantly, it would provide information on which some action can be taken that would be aimed at improving the education received by the students.

The Board is not in favor of comparisons for comparison's sake, and has adopted policies for reporting State results that would avoid the inclination to misuse or misinterpret results through simple ranking of State scores.

1. Authorizing State-representative assessments.

Congress authorized NAEP trial State assessments in eighth grade mathematics in 1990 and in fourth grade reading and fourth and eighth grade mathematics in 1992. No subsequent State testing is authorized. These trial State assessments are being closely monitored by the Department of Education, the NAEP contractor and the Board and will be the subject of analysis across the country. A congressionally required independent evaluation is being conducted by the National Academy of Education; reports are expected in March and September 1991 and in late 1993.

On June 6, 1991, the National Center for Education Statistics will release data from the 1990 State trial assessment in eighth grade mathematics, along with national results for grades four, eight and twelve. Almost forty States have already signed up for the 1992 State trial assessments and there are likely to be more. Unless the evaluation proves that the State trials do not yield valid, fair and accurate data, and unless the State interest in participating evaporates, the groundswell of support for regular State-representative data as a part of NAEP is likely to continue.

The Board believes that, given favorable results from the congressionally required evaluation, the NAEP legislation should be amended to provide a general authority to conduct State representative assessments, with the frequency and subjects determined according to Board policy, with the full costs borne by the federal government and with State participation voluntary.

2. Restoring local options to use NAEP

The Board believes that the NAEP legislation should be amended to permit States, school districts and non-public schools to use NAEP, at their option and cost, to collect data and report on performance of schools and school districts. This would restore the option that was eliminated under the 1988 reauthorization of NAEP.

It is not the policy of the Board that reporting below the State level should be a federal activity. The Board has never suggested-

-nor does it now--that the organization that conducts NAEP should establish samples in individual schools and school districts as part of the NAEP contract and report the results. The Board views this as a local control issue--just as no State or district should be required to participate in NAEP, neither should a State or district be prevented from using NAEP to assess and report student performance in its schools.

3. Conducting NAEP on an annual rather than biennial schedule.

The Board believes it would be better to give NAEP on a rotating schedule of three subjects per year rather than the current practice of testing three to five subjects every two years. Annual NAEP testing and reporting are necessary to provide timely and sufficient data for policy-makers and the public. This would allow NAEP to provide key measures of academic achievement for the annual progress reports on American education, called for last year at the Education Summit between President Bush and the nation's Governors.

An annual NAEP would produce little net increase in test burden because the current every-other-year cycle includes up to five tests each testing year and extensive pre-testing in off-years. NAEP conducted annual testing from 1969 until 1980, so this proposal is not without precedent.

The Board believes a permissive authority allowing annual testing rather than a mandated biennial cycle would provide the flexibility for NAEP to be responsive to information needs as they evolve. Of course, the number of subjects actually assessed would be subject to the availability of appropriations.

4. Clarifying the role of the Board.

The Board believes that, with respect to its role and responsibilities, the NAEP legislation contains ambiguities and conflicting provisions. For example, the NAEP legislation gives the Board the authority to "formulate the policy guidelines for the National Assessment," to "exercise its functions, powers and duties independent of the Secretary and the other offices and officers of the Department of Education," to "exercise its independent judgment, free from inappropriate influences and special interests," and to "hire its own staff."

Taken together, these provisions could not be more clear in their intent to establish an independent policy body for NAEP insulated from the jurisdiction of the Department of Education. However, other provisions reduce this clarity.

Under the NAEP legislation, the Commissioner of Education Statistics carries out NAEP "[w]ith the advice of the National Assessment Governing Board [emphasis added]." This has been interpreted by the Department of Education to mean that the Board is an advisory committee whose policy decisions must only be given consideration by the Secretary. Another part of the law provides that "The Secretary may appoint, at the direction of the Board" up to six technical employees under an excepted service appointing authority. This provision has the potential for eroding the Board's independence and contradicts the provision that the Board shall hire its own staff.

Another issue relates to the Board's budget, which as a set-aside within the NAEP line item in the Department's Assessment, Statistics, Research and Improvement appropriation, is in fact the Secretary's responsibility to administer. The contradiction is painfully obvious--independence from the Department is incompatible with the Secretary's budget responsibility.

While I have highlighted a few of the areas of concern with respect to ambiguities in the role of the Board, attachment A contains a more complete analysis. However, I would point out that while we have not always agreed, for the most part, these issues have been handled openly and with a general spirit of cooperation with the Department.

But these issues are fundamental. Their existence erodes the principles in the law on which the Board was established: independence from the Department of Education, tripartite check and balance NAEP governance structure, and freedom from inappropriate influences and special interests. The Board's only goal in this analysis is no more than what it believes the Congress intended--a truly independent Board whose policies direct the conduct of NAEP.

Thank you very much; I would be pleased to answer any questions you may have.

Attachment A: Policies Adopted by the Board Requiring Legislative Action

Attachment B: Statement on the Future of NAEP

ATTACHMENT A

THE NATION'S REPORT CARD naep

National Assessment Governing Board

National Assessment of Educational Progress

Policies Adopted by the
National Assessment Governing Board
That Would Require Legislative Action

January, 1991

1100 L Street, N.W.
Suite 7322
Washington, D.C. 20005-4013
(202) 357-6938

BACKGROUND

During the past year, the National Assessment Governing Board adopted two comprehensive policy statements. In December 1989, the Board adopted positions on the Future of the National Assessment addressing nine specific policy issues. In November 1990, the Board adopted a statement on policy issues for the 1994-96 NAEP contract award addressing twenty-four specific policy issues. Some, but not all, of these policies cannot be implemented without amending the current NAEP authorization. This report has been prepared with the purpose of highlighting these policies.

In all cases, the policies that follow have been stated in a format that makes clear what is being recommended for legislation and may not be a word-for-word restatement of the original policy.

STATE-REPRESENTATIVE ASSESSMENTS

1. **Provide a general authority to conduct State-representative assessments, with the frequency and subjects determined according to Board policy, with the full costs borne by the federal government, and with State participation voluntary.**

Rationale

During 1990, both the assessment committee of the Council of Chief State School Officers and the National Forum on Educational Statistics have called for regular NAEP state-by-state assessment. The need for these data has been emphasized in an analysis of national education goals recently prepared by the Urban Institute. Also, State NAEP samples clearly are needed by the President and the Governors for their annual progress report on American education.

Without testing on a common instrument under common conditions it would be very difficult for citizens of any State to have sound comparative information on how much their students have learned. SATs and ACTs are taken by self-selected groups. Commercial standardized tests have their "Lake Woebegone effect" in which most children appear to be above average. State criterion-referenced tests vary widely.

Congress authorized State NAEP on a trial basis in 8th grade math in 1990 and in 4th and 8th grade math plus 4th grade reading in 1992. No subsequent State testing is authorized. These trial State assessments are being closely monitored by the NAEP contractor and will be the subject of analysis across the country. An independent evaluation, required by law, is being conducted by the National Academy of Education; reports are due in late 1991 and 1993.

If these trial assessments are successful and the information useful, the support will likely be strong to proceed with State-based assessments in the NAEP program. If there are problems, changes in the assessments should be made. The underlying reason for State assessments is perhaps stronger today than ever and that is the need for sound, comparative data.

The Board believes that full federal funding is needed to help move to a testing program in which NAEP data are collected for every State, that there is a clear public interest in collecting such data, and that it is the only way a national progress report on American education can be complete.

Federal funding for many years has supported the collection and reporting of education data by State. These are essentially limited to "input" data including: enrollment, revenues, expenditures, program participation of students, pupil-teacher ratios, teacher characteristics, and high school graduation requirements. There is no dispute that a federal interest exists for the collection and reporting of such information at full cost to the government.

However, some analysts argue that there is only a very limited federal interest in collecting and reporting the outcomes of the application of these resources and requirements by State, that the outcome of the education process is a State concern and, therefore, State participation in NAEP should be limited to those voluntarily willing to assume part of the costs.

This view is perplexing. Almost no valid information exists about education outcomes at levels of analysis that bear on national and State policymaking. Policymakers at all levels are increasingly demanding such information and only the federal government is positioned, through NAEP, to provide it. Additionally, the amount of money involved is not excessive, particularly in comparison to other federal education initiatives.

The fact alone that President Bush and the Governors have, for the first time in our nation's history, set national education goals for the United States represents a sea change in the partnership in education between the federal and State governments. The fact that these goals describe outcomes underscores the concern for results. Responsibility for education is first of all a State matter; and on this basis States will be expected to do the "heavy lifting" in effecting reforms and improving education performance. It seems, therefore, fair and appropriate that the Federal government, in fulfilling its role in the partnership, provide resources needed to help assess the results of State education improvement efforts.

Finally, while provisions related to voluntary costs to the States should be removed, those related to voluntary participation in the assessment should not. The intent is to encourage full State participation in NAEP by removing a federally wrought impediment; under no circumstance should participation by a State in NAEP be mandated, constitute a requirement for receiving any federal benefit, or in any way be other than voluntary.

ANNUAL NAEP ASSESSMENTS

2. **Provide for annual rather than biennial assessments in the National Assessment of Educational Progress.**

Rationale

It would be better to give NAEP on a rotating schedule of three subjects per year instead of maintaining the current practice of testing three to five subjects every two years.

Annual NAEP testing and reporting are necessary to provide timely and sufficient data for policy-makers and the public. This would allow NAEP to provide key measures of academic achievement for the annual progress reports on American education, called for last year at the Education Summit between President Bush and the nation's Governors.

For example, reading and mathematics plus one other subject could be tested in even-numbered years. Science and two other subjects could be tested in odd-numbered years. Annual schedules for data collection and reporting would produce cost savings by stabilizing work-flow and eliminating separate pretest samples for new items.

An annual NAEP would produce little net increase in test burden because the current every-other-year cycle includes up to five tests each testing year and extensive pretesting in off-years. Also, testing in three subjects annually would provide ample opportunity on a scheduled basis, perhaps every six years or so, to test in such subject areas as foreign languages, the arts, and economics, which now could only be assessed sporadically. The Board already is considering these subjects for possible assessments in 1996. It is important to test a range of subjects to discourage any narrowing of the curriculum.

NAEP conducted annual testing from 1969 until 1980 when testing was placed on a biennial schedule to save funds, so this proposal is not without precedent. Amending the current authority to permit annual testing and reporting rather than mandating a biennial cycle will provide more flexibility in the conduct of NAEP and allow it to be more responsive as information needs evolve. Of course, the number of subjects actually assessed would be subject to the availability of appropriations.

RESTORING LOCAL OPTIONS TO USE NAEP

3. Eliminate the prohibition against the use of NAEP test items and reporting below the State level by States, school districts, and non-public schools and school organizations. Modify the NAEP confidentiality provision to permit States, school districts, and non-public schools and school organizations to use NAEP for providing results on individual districts, schools and students. The cost of such uses of NAEP and of reporting would be paid by the requesting agency. Reporting of individual student scores would continue to be prohibited.

Rationale

These changes are related. They would restore the local option to use NAEP that States, localities, and schools previously had prior to 1988. The Governing Board has never suggested--nor does it now--that the organization which conducts NAEP should establish samples in individual schools and school districts as part of the regular NAEP contract and report the results. However, it is highly desirable that States, school districts, and schools have the option--if they wish to pay for it--of using NAEP to measure their own schools, as some have done in the past.

During the 1970s and 1980s a number of States used NAEP items and even NAEP tests to gather information on school and school district performance. NAEP tests and test items also had been used, apart from the regular assessment, for individual student testing in many State and local testing programs. All this activity was a desirable effort to provide useful data to students, parents, and schools.

However, in 1988 Congress enacted two provisions under NAEP that prevent State and local use of NAEP. The first prohibits any use below the State level of NAEP "test items and test data employed in the [State NAEP] pilot program." Although the statutory prohibition pertains only to questions used in the state-by-state assessments of 1990 and 1992, NCES has applied it to all NAEP exercises starting in 1990. The second provision requires the Commissioner to ensure "that all personally identifiable information about students, their educational performance, and their families and that information with respect to individual schools remain confidential..."

In July 1990, the National Forum on Educational Statistics, a group of 50 State representatives and federal agency officials convened by NCES, urged that States, "if they wish, [should be able to] analyze...student achievement [on NAEP]...so that comparisons could be made among education units by significant subgroups."

Even though supporters say the prohibition and the confidentiality provision protect local interests and family privacy, they also have an opposite effect--of denying the previously available option of using NAEP if localities wish to do so. The Board believes that these two provisions should be amended to clarify that they apply to the federal government and its contractors in the conduct of NAEP; to permit State and local use and augmentation of NAEP at local option and cost and pursuant to procedures established by the Commissioner to ensure test security, uniform administration and valid reporting; and to prohibit public reporting of individual student scores.

Large-scale use of NAEP raises serious issues of test security and uniform test administration and reporting. If NAEP tests are purchased by schools, given without supervision by thousands of classroom teachers, and reused for many years--as is done now with commercial tests--there is serious danger local results will be inflated and comparisons with national NAEP norms rendered invalid. Because of differences in motivation and possible preparation there would also be a serious problem in comparing the results of students who took NAEP knowing they would get individual scores with those in the national and State samples who are assured NAEP can't "count" for them. Also, the design of each NAEP subject area assessment--now divided into blocks for matrix sampling--would have to be modified to provide for individual student testing and reporting.

These concerns, however, can be addressed and should not be the basis for denying the use of NAEP and NAEP items by States and local education entities.

INTERNATIONAL COMPARISONS

4. Provide authority for a regular international component for NAEP.

Rationale

NAEP should have a regular international component funded through its regular appropriation. The purpose of this component would not be assessment hegemony, i.e. transforming NAEP into the primary instrument by which participating nations compare their performance. Instead, the purpose of an international component in NAEP would be to provide information to help interpret and understand NAEP results, just as State-representative data are intended to enhance understanding of the performance of State education systems.

The President and the Governors have set national education goals in an international perspective (e.g. By the year 2000, United States students will be first in the world in mathematics and science) because a fuller understanding of the performance of U.S. students must take into account the education performance of our neighbors, trading partners and competitors.

Making valid international comparisons is a very complex undertaking. Curriculum and instructional practices, language and culture, student composition, provisions for special education and disadvantaged students, rate of high school completions, and the composition of in-school age groups all vary widely from country to country and prevent easy comparisons. However, if we are to have effective benchmarks for our own education practices and performance, they must be informed by expectations held for comparable students in other countries and the aggregate performance of those students.

Current law charges NAGB with "identifying appropriate achievement goals for each age and grade in each subject area tested under [NAEP]," "developing standards for interstate, regional and national comparisons," and "taking appropriate actions to improve the form and use of [NAEP]." Faithfully fulfilling these obligations demands not merely describing what U.S. students know and can do, but determining whether it is good enough. Making this determination should involve appropriate analysis, in all subjects where practicable, of the performance of students in other nations.

For example, in setting achievement goals, it is essential to assure that they are both challenging and reasonable; examining the curricula of other countries would help in making this determination. Where other countries have comparable curricula in a particular subject area and a comparable student sample could be drawn, it would be useful to analyze respective patterns of student performance. It is not likely that this would involve numerous countries for each assessment, and it is not intended that such efforts would amount to a true international assessment. Information collected under the international component would be directed solely at informing U.S. results.

In 1988 an international comparison, using NAEP items in mathematics and science, was conducted by Educational Testing Service under grants from the National Science Foundation and the U.S. Department of Education. A similar comparative assessment, with similar funding, is planned by ETS in 1991. It will use items developed jointly by participating countries, including some from NAEP. Also planned are international assessments in science and mathematics in 1994 and 1998 to be conducted by the IEA, the timing of which coincides with the likely NAEP schedule for these subjects and which may be linked with NAEP. While the Board applauds these efforts, the fact is that they are conducted outside the regular NAEP framework. Thus, they are not a dependable source of information for developing policy (e.g., setting subject area achievement levels) and are not subject to NAGB policy.

With authority for an international component, NAEP could do a better and more complete job of reporting to the American people on the performance of American students. Having specific legislative authority for this purpose is essential to avoid reliance on data intended for other purposes, to assure orderly and systematic planning, preparation and data gathering, and to coordinate policy development with other agencies conducting international assessments.

ASSESS RANGE OF SKILLS

5. Modify the NAEP statement of purpose to clarify that assessment will cover the broad range of skills and knowledge (not just the basic skills) in the subject areas to be assessed.

Rationale

This clarifying amendment should be non-controversial. It reflects the view of subject matter and testing experts, expressed through the consensus process and in public hearings we have held, that NAEP assessments should include an examination of the higher order thinking skills of U.S. students as well as "basic skills." Our policy and practice already incorporate this view, but we believe specifying it in the law would be useful.

REGIONAL SAMPLES

6. Replace the requirement to employ regionally-representative samples and produce regionally-representative data with language that would make doing so a permissive activity.

Rationale

Regional reporting was from the start of NAEP intended as a substitute for State reporting. From a policy perspective, there is little cause for action that can be derived from such data, and there is little evidence that such reporting has had any effect on education policy.

With the possible advent of regular, predictable State level data collection, the need for drawing regional samples declines. While we do not advocate prohibiting regional samples and reporting, we do envision a time in the future when such sampling and reporting may be unnecessary. Thus, the Board should have the flexibility to recommend its application in the future consistent with other changes in NAEP.

INDEPENDENCE OF NAGB

7. Eliminate ambiguities and conflicting provisions in the NAEP legislation regarding the role and responsibilities of the National Assessment Governing Board (NAGB).

Rationale

The NAEP legislation gives NAGB the authority to "formulate the policy guidelines for the National Assessment," to "exercise its functions, powers and duties...independent of the Secretary and the other offices and officers of the Department of Education," to "exercise its independent judgment, free from inappropriate influences and special interests," and to "hire its own staff."

Taken together, these provisions could not be more clear in their intent to establish an independent policy body for NAEP insulated from the jurisdiction of the Department of Education. However, other provisions reduce this clarity.

Under the NAEP legislation, the Commissioner of Education Statistics carries out NAEP "[w]ith the advice of the National Assessment Governing Board [emphasis added]." Further, the law provides that NAGB "shall have the authorities authorized by the Federal Advisory Committee Act (FACA) and shall be subject to the open meeting provisions of that law." These provisions have been interpreted by the Department to mean that NAGB is an advisory committee subject to Department policies and procedures for advisory committees and that NAGB policy decisions must only be given consideration by the Secretary. Although the Department now agrees that NAGB is not an advisory committee, it continues to oversee NAGB activities as if it were, due in large part to the reference to FACA. Still unresolved is the degree to which the Department is obliged to observe policy established for NAEP by NAGB, rather than merely to consider it as advice.

Another issue relates to the NAGB budget, which, as a set-aside within the NAEP line item in the Department's Research, Statistics and Improvement appropriation, is in fact the Secretary's responsibility to administer. The contradiction is painfully obvious--independence from the Department is incompatible with the Secretary's budget responsibility. So far this issue has been managed in a spirit of cooperation, but it remains a potentially fatal flaw in NAEP's governance design.

There is a corollary to the budget issue. The Department has asserted that the Secretary's responsibility for administering and monitoring Department appropriations brings NAGB activities under its jurisdiction. Where compliance with government-wide requirements for fiscal accountability are at issue there is no disagreement, but assertions by the Department that its fiscal authority under the current arrangement extends to review and clearance of the intellectual activity of the Board and its staff in conducting its work is incompatible with NAGB's statutory mission and is simply unacceptable.

Another part of the law provides that "The Secretary may appoint, at the direction of the Board" up to six technical employees under an excepted service appointing authority. Once again, while managed in a spirit of cooperation, this provision has the potential for eroding NAGB independence and contradicts the provision that the Board shall hire its own staff.

An area of ambiguity relates to evaluation of NAEP. Evaluation, particularly as prescribed under the NAEP legislation, is intended to bear on policy formulation. A commonly accepted standard for objectivity involves insulation of the evaluator from the subject of the evaluation. The NAEP legislation places responsibility for evaluation of NAEP with the Commissioner and is silent about the evaluation responsibility of NAGB. We believe that this is the opposite of sound practice for assuring objective evaluations and would advocate for an amendment that would add to NAGB's functions the formulation of policy for the conduct of evaluations of NAEP.

While the law is very specific regarding the balance of responsibilities between the Secretary and the Board in making appointments to the Board, it is silent with respect to how NAGB will organize itself in performing its functions. The Board believes that the independence provisions in the law authorize total autonomy with respect to its internal workings. The Department believes this to be true with one exception--the appointment of the Board Chairman by the Secretary. Although this issue has been managed in a spirit of good will and cooperation, it remains a major point of contention.

These issues are fundamental. Their existence erodes the principles on which NAGB was established: independence from the Department of Education, tripartite check and balance NAEP governance structure, and freedom from inappropriate influences and special interests.

The goal that the Board seeks is no more than what Congress intended--a truly independent body whose policies direct the conduct of NAEP. The NAEP legislation does not define independence, but the Board believes that it includes three factors: (a) freedom from inappropriate influences and special interests in making administrative and policy decisions; (b) assurance that policy decisions made by NAGB will be implemented in the conduct of NAEP; and (c) full, easy and timely access to information that NAGB needs from a variety of sources, including the Department, the Congress, education practitioners, and the NAEP contractor.

Legislation designed to achieve these ends could take many forms, from modifying or eliminating some of the provisions cited above to establishing NAGB with its current authorities as an independent agency outside of the Department. Should the Department of Education be contemplating changes in NAGB's authority as part of its reauthorization proposal, the Board would look forward to discussing this matter.

NAGB MEMBERSHIP

8. Provide for the continuing membership of the present Board according to current terms and for continuing the current provisions for filling vacancies. Add a provision that explicitly limits Board members to two consecutive terms of up to four years in each term. Encourage the Secretary to consider a candidate's previous experience on the Board and the overall experience of the Board when making appointments.

Rationale

The current legislation contains provisions for the transition in governance from the Assessment Policy Committee to NAGB. These provisions are now inapplicable and will be removed whenever NAEP is reauthorized; new provisions should leave no doubt about the continuation of the current membership.

During one thirteen month period in 1989-90, 14 of the 23 appointed members of the Board were replaced. This turnover rate is too high to assure the stability and continuity that was envisioned in providing for terms of up to four years. Providing for two terms and taking Board experience into account in making appointments will help assure stability.

ATTACHMENT B

National Assessment Governing Board

National Assessment of Educational Progress

NATIONAL ASSESSMENT GOVERNING BOARD

Positions on
The Future of the National Assessment

APPROVED: December 9, 1989
At Meeting in Austin, Texas

The National Assessment Governing Board, mindful of its statutory responsibility to seek to improve the National Assessment of Educational Progress, hereby adopts these positions and recommendations:

1. The National Assessment of Educational Progress should provide information for an annual report card by testing at least three subjects each year.

The exact configuration should be determined after a careful analysis of cost and management considerations involved in such an increase over the current every-other-year testing cycle.

Rationale

Annual NAEP testing and reporting are necessary to provide timely and sufficient data for policy makers and the public. Under its current schedule, however, the National Assessment is woefully incomplete. Reporting assessment data each year will allow NAEP to become the key measure of academic achievement in the annual Report Cards on American schools, called for in September at the education summit between President Bush and the nation's governors. Also, regular annual data are needed from NAEP to replace the Education Department's annual "wall chart" which relies on SAT and ACT scores.

1100 L Street, N.W.
Suite 7322
Washington, D.C. 20005-4013
(202) 357-6938

2. NAEP should move as quickly as feasible to full state participation in all subjects and all three grade levels (4th, 8th, and 12th) tested. No state, however, should be compelled to participate. The federal government should pay the full cost of the state-by-state NAEP program.

In 1990 each state must pay about $100,000 to take part in the state-by-state mathematics assessment. Several have cited this expense as a reason for not participating.

Rationale

The Governing Board wishes to have a full testing program in which NAEP data are collected for every state. There is a clear public interest in obtaining such data; it is the only way a national Report Card can be complete. The Board believes it would be inadvisable to require states to participate in NAEP. However, full federal funding would make it less likely that states would decline to take part. By having NAEP testing in all states, the Board also wishes to end duplicate national samples as soon as feasible.

3. The Governing Board urges Congress to remove the prohibition against the use of NAEP tests and data reporting below the state level.

Publication of individual student scores should continue to be prohibited. Any additional costs for testing below the state-level samples should be paid by participating states or localities.

Rationale

The clause to be eliminated--Subsection (4)(C) of P.L. 100-297--reads as follows: "The use of National Assessment test items and test data employed in the pilot program authorized in subsection (2)(C) to rank, compare, or otherwise evaluate individual students, schools, or school districts is prohibited."

This change would permit NAEP test items and data to be used at the levels--school district and school building--where many important educational decisions are made. It would provide helpful information to parents and others interested in school improvement. By making NAEP more useful to local schools, offering this option would encourage them to participate in national and state sampling. States and localities would be able to tie their own regular testing programs into NAEP. Before the prohibition was enacted in 1988, many schools and districts requested NAEP reports on students who participated in national sampling; several states administered NAEP items.

4. NAEP should establish international samples which would participate regularly in the Assessment program.

Other nations should be invited to participate in NAEP on a regular schedule. NAEP tests should be administered to representative samples of their students, which might serve as reference points for achievement in American schools.

Rationale

The Report Card for American students is incomplete if it does not include a regular barometer of academic achievement in an international context. Today there is no such barometer.

5. The Governing Board believes the release of NAEP data must be speeded up. It requests its staff to report by March on what steps could be taken to reduce the time for reporting NAEP test results, including possible design changes if needed.

It has taken 21 to 24 months after testing to report NAEP results. This is scheduled to improve to 15 to 18 months for the 1990 assessment. But NAEP will still be much slower than the College Board and American College Testing Program summary reports and almost all state testing programs, which usually make data public in less than six months.

Rationale

The long delay in reporting severely reduces NAEP's usefulness and impact. It will be difficult for NAEP to serve as an annual Report Card if it is always more than a year late.

Money is not the cause of the lengthy delay. Rather, it stems from NAEP's highly complicated survey design and several basic decisions about how it is conducted. These issues should be studied carefully, including any trade-offs that might be necessary if NAEP design, test construction, and testing practices are changed. The study will include a cost-benefit analysis of any recommendations.

6. The Governing Board believes the time needed to develop NAEP exams must be reduced. It requests its staff to report by March on what steps could be taken to shorten the time for planning and developing NAEP exams.

Each new test now takes about 30 months to develop. This includes the consensus process for goals and objectives, preparation of test specifications and questions, review of questions by NAGB, and reviews of cognitive items and background questions by the Office of Management and Budget and the Education Department.

Rationale

The study should include all aspects of the test preparation cycle. Special examination should be made of the impact on NAEP of the Paperwork Reduction Act of 1980; reviews by OMB and the Education Department presently take a total of eight months per cycle. The study will include a cost-benefit analysis of any recommendations.

7. NAEP exams should be revised to include the full range of knowledge and skills from basic skills to advanced subject-matter knowledge and analytical, integrative skills.

Test planning committees and item writers should include a significant number of test objectives and questions on NAEP exams that assess higher-level analytical skills and knowledge as well as basic skills.

Rationale

Attention to higher-level skills and knowledge as well as basic skills is clearly in line with efforts to strengthen the school curriculum. It will permit NAGB to establish high standards when it sets goals for NAEP tests. The NAEP math exam has already been changed in this direction.

To achieve this goal NAEP exams may have to include more open-ended questions and essays instead of relying as heavily as they do on multiple-choice items. The multiple-choice questions themselves may also be changed to test more higher-level skills and knowledge than they do now.

8. NAEP should obtain data on additional important groups.

Sample sizes should be enlarged to provide information on groups such as low-income children and those attending private schools.

Rationale

This would provide important information for educators and policy makers and for analysis of schools. The definition of low-income students may be difficult, but NAEP could use those who qualify for participation in federal free or reduced-price lunch programs or children who attend Chapter I schools.

9. Change in the governance structure of NAEP should be considered by Congress.

The role and responsibilities of the Governing Board should be clarified by new legislation.

Rationale

The current system of divided authority may lead to confusion and disputes which would hamper the NAEP program. Governance and administrative supervision of the National Assessment should be rationalized and focussed. Overlap and duplication of functions of NAGB and NCES staffs should be eliminated.

Chairman **Kildee**. Thank you very much. We will go through the panel then address our questions singularly or jointly at the conclusion.

Dr. **Koretz**. Mr. Chairman, thank you for the opportunity to address the subcommittee about the National Assessment and about national testing issues in general.

I also will abbreviate my remarks and ask that my full testimony be submitted to the record.

Chairman **Kildee**. Without objection, it will be included.

Dr. **Koretz**. I am going to start with a couple of general issues about national testing because those are the underpinnings for the specific comments I will make about the National Assessment.

I think to judge how well the National Assessment and other tests can meet the rapidly rising expectations people in the policy world have for them, it is necessary to bear three issues in mind.

The first is the multiple functions that these tests are being asked to serve. The second is the problem of corruptibility of test scores, and the third is the unavoidable technical complexity of the whole enterprise.

Current proposals ask the NAEP to do three things: to monitor achievement which is its traditional function, to describe what our kids can do; second, to evaluate educational programs or policies; and third, I think underneath it all to hold people accountable for performance.

The pell mell rush to increase testing that the letter that Mr. Goodling read referred to is not a pell mell rush for more monitoring. It is a pell mell rush to judgment. It is, I think, currently reflective of a desire for more evaluation and more accountability.

Unfortunately, it is very difficult, some would say impossible, for a single test to serve all three of those functions at the same time. There are two reasons for this. I can come back to them in more detail later if you would like or in writing. The first is that tests have to be designed differently for those different functions, and the second, which I think is perhaps more important, is that using a test for accountability can undermine its accuracy as a monitoring tool, as a source of good, descriptive information about what our kids can do.

And this points to the second of my general issues which is the corruptibility of tests. In the policy world there is a widespread view that tests are kind of like a fever thermometer, a measure that you can always trust. If a fever thermometer shows a fever going down, you can be confident that the fever has gone down. I don't think that is necessarily the case, and, in fact, we are surrounded by measures that are not so incorruptible.

One example might be airline on-time statistics. If you suddenly find that you are getting someplace on time, is it really because performance has improved, or is it because the formal length of the flight has been increased to allow for the same delays on the tarmac? If it is the latter, which I think in some cases it has been, it doesn't tell you anything about how long it gets to where you want to go.

In the same way tests can be corrupted if they are used for accountability. Scores can go up, but actual achievement, the time it takes to get to Detroit, may not change a whit, and this is the basis, one basis, for what we now call the Lake Woebegone phenomenon, the puzzling facts that most States and districts claim to be above average.

One reason this can happen is very simple, and it was also referred to in the letter that Mr. Goodling read, which is teaching to the test. Teaching to the test focuses instruction on the narrow content of the test rather than the much broader range of achievement that the test really is supposed to inform us about.

The final general point is that large scale assessments such as NAEP, if they are to provide useful information that we can trust, are technically highly complex. Currently there is a very widespread impatience in the policy community with technical aspects of measurement. They are seen in many quarters as kind of an irritating obstacle to changes that people want to make, but that impatience doesn't in any way decrease their importance.

With those basic points in mind, let me turn to the NAEP itself, the National Assessment.

In my view the National Assessment remains ideally suited to its original function of monitoring what our kids can do. Moreover, there is at present nothing else that we have that is a close substitute for it. Nothing else tests nationally representative samples of kids at frequent intervals on a wide variety of subjects, and in addition the NAEP, unlike many State level tests, is not yet taught to in a way that gives us reason to distrust its results.

So if we lose the NAEP's ability to monitor achievement, we have lost a lot.

The National Assessment could be improved in this regard. It is not to say it is perfect for monitoring. Some of the potential improvements, however, would cost money, and money is scarce not only because of the general budgetary situation but because a lot of money is going into assessment goals that I think are quite futile such as using State comparisons to evaluate State educational programs.

I won't take time to go into what some of the improvements might be, but I will just mention one. The Congress has been interested for at least eight or nine years in the performance of high achieving students in the pipeline, future scientists and mathematicians. The National Assessment is not designed well to assess the performance of those students. We don't have a trustworthy indicator of the performance of those students. We could make the National Assessment into one if we wanted to spend the money to do it.

The second function, evaluation, is really the key to the argument about State comparisons using NAEP. If you read the justifications for State comparisons, it seems to be very clear nobody is really very interested in simple, descriptive information that says that mathematics achievement is not as good in Mississippi as it is in Minnesota.

What people want, and Dr. Walberg alluded to this, is some indication of which States have better educational programs or policies. The fly in the ointment is that there is simply no way that the National Assessment as it is currently structured can provide meaningful information of that sort. When one State outscores another, it might be for educational reasons or it might not, and we won't know when it is.

And moreover when State A does outscore State B for educational reasons, we won't be able to tell which educational reasons.

The same ambiguities and uncertainties would arise if the National Assessment were used at the local level if the statute were changed to allow that.

The reasons for this are too complex to discuss in detail here. I would be happy to come back to them, but they really boil down to one simple point which is to say that a policy causes educational achievement to go up or down, you have to be able to rule out reasonable alternative explanations of the sort that, Mr. Goodling, you mentioned in your letter.

The National Assessment is not designed to do that. It doesn't track the growth of students over time. It includes only very weak information on the background and characteristics of students or their communities, and it doesn't include much information on the educational policies or practices in the districts where students go to school.

Some people say that is all right. It is still adequate for some degree of evaluation of policies and programs, but in saying that, they are implicitly accepting a standard of evidence that is vastly looser than we would be willing to accept in any other domain. We wouldn't, for instance, in evaluating health programs or medical treatments, be willing to accept evaluations based on evidence that is flimsy. If we did, what we would have is incorrect evaluations that would sometimes mean that effective medical treatments go down the tubes and would sometimes mean that ineffective medical treatments would continue. The same will happen in education if we accept a low standard of evidence.

The National Assessment could be redesigned to provide this function. There is no reason why it couldn't. It would require a wholesale redesign of the program. It would cost a lot of money, and absent a lot of additional money it would therefore take away from the things that are currently appropriately spent to increase the quality of the descriptive information that the National Assessment is designed to give us.

The third function, which is really very tidily tied to the second, is accountability. Using the National Assessment or another national test for accountability purposes would in part run into the same difficulties I have just described because it isn't really very reasonable to hold people accountable for differences in educational performance unless you can really say that they are responsible for them.

But there is another reason to be worried about accountability beyond that which is that using the National Assessment in that way could undermine its value as our only high quality nationally representative indicator of performance.

The issue, as I suggested earlier, is teaching to the test. In the educational community and in the policy community there is a lot of disagreement about whether teaching to the test is, from an instructional point of view, desirable or pernicious. I think for present purposes we don't have to be distracted by that argument.

I think, by the way, it is largely pernicious, but that is beside the point.

The point is regardless of whether it is desirable or pernicious, it can wreak havoc with the information, the descriptive information, we want from a monitoring system.

The National Assessment does include some protections against that kind of corruption. One is fairly rigorous test security. The second, which Dr. Walberg mentioned, is a sampling system where each kid takes only a fraction of the test. It is by no means clear to me that those are enough to protect it from corruption.

One last point on this is that I think local comparisons are by and large riskier than State comparisons in this regard because they increase both the incentive to raise scores and the tools which people can raise scores.

Proponents of local comparison say, "Exactly. That is what we want people to do. We want them to increase their performance." Regardless of whether that is a desirable thing from an educational point of view, it is not a desirable thing if you want the National Assessment to tell you what kids can do.

There is a solution to this dilemma which is to use the National Assessment for monitoring and to use something else for accountability if test based accountability is what you want to do.

Finally, I would just like to make a couple of comments about the National Assessment Governing Board in the context of my third point which is the unavoidable technical complexity of the National Assessment.

Under the current statute the National Assessment Governing Board has an extremely broad purview. It includes policy questions that don't require a great deal of technical expertise and that probably should not be considered the province of technical experts.

It also, however, includes explicitly in the statute authority over very technically complex issues such as designing the methodology of the assessment.

The National Assessment Governing Board is simply not constituted appropriately for the second category of functions. Right now only two of the 23 seats on the committee, on the board, are set aside for experts in testing and measurement, unlike, say, the Educational Testing Service, which is the prime contractor for the NAEP, does not have a standing panel of experts to advise it.

The contrast with other government funded agencies or organizations that are responsible for making technical decisions, an example I would use is National Institutes of Health or ADAMHA review panels. It is very striking.

We would not constitute a panel under NIH's purview to design clinical trials that comprises primarily people who don't know how to design clinical trials; but that is precisely what we have done with the National Assessment Governing Board.

This lack of expertise is important for several reasons. One is, as I suggested before, there really is no way to separate technical from substantive considerations in testing. In medicine technical considerations tell you whether a medicine works. In testing technical considerations tell you whether you can believe the information you get or alternatively what it means.

Second, there is extraordinary political pressure now from various education constituencies to change the assessment, to use it in many new ways and to get moving with these changes. The National Assessment Governing Board needs the expertise to understand when those changes are practical and when not and to decide what they have to do to accommodate those pressures in a technically defensible manner.

Therefore, I think the effectiveness of the National Assessment could be improved by changing the statute to insure reasonable representation of technical expertise on the governing board. How much change is required, I think, depends on whether or not the purview of the board remains as it is.

If the board does retain authority over highly complex issues, I would suggest that it might be reasonable to increase the representation of technical experts to as much as a third of the board which would be eight members, currently.

If authority over some entirely technical issues were removed from the board's purview or made joint purview of the board and some other agency, perhaps a more modest change would make sense such as increasing the representation to four.

In either case I think a requirement should be put in place that the individuals chosen for those slots are widely recognized as experts by their professional peers throughout the United States. There is currently no such requirement. Again, I think that NIH and ADAMHA review panels might provide us with some hints about how we could assure that in the future.

Finally, to wrap up since I have gone on a bit to long, clearly we are at a turning point, not just for the National Assessment but for educational assessment and assessment at the national level in general. The National Assessment has already, in my view, taken serious States in its trial State assessment away from its traditional monitoring function and toward the functions of evaluation and accountability, and I think many of the proposals on the table now, starting with the last session, would push it further in that direction.

My own view is that the changes made to date were undesirable and that further changes in those directions would be ill advised. Absent dramatic changes in the structure of the NAEP, efforts to use it for evaluation are futile and potentially misleading, and efforts to use it for accountability will simply undermine its value as our only sound indicator of the performance of our nation's youth.

You may well decide that the monitoring function is not, in your view, the most important function for the National Assessment. There are certainly many people who don't believe that it is the most important function. But as you deliberate that, I would ask you to bear in mind that although you will hear claims to the contrary, we really can't have our cake and eat it, too. We can't have everything from the National Assessment. It either is going to do what it was designed to do or it is going to do something different, and we have to make a choice. My hope is that we will choose to continue using it and improving it as a tool for monitoring performance.

Mr. Chairman, thank you again for the chance to address the subcommittee. I would be happy to take questions at any time.

[The prepared statement of Daniel M. Koretz follows:]

STATEMENT OF

Daniel M. Koretz
Senior Social Scientist
The RAND Corporation[1]

before the

Subcommittee on Elementary, Secondary, and Vocational Education
Committee on Education and Labor
U.S. House of Representatives

March 13, 1991

---

[1] The views expressed here are solely those of the author. They do not necessarily represent the position of the RAND Corporation or its clients and were not prepared in fulfillment of RAND's contracts or grants.

Mr. Chairman, thank you for the opportunity to address the Subcommittee about the National Assessment of Educational Progress (NAEP). Before turning specifically to the operation of and statutory basis for the NAEP, I will briefly address several general issues about testing that underlie my specific recommendations for the NAEP.

BACKGROUND

The NAEP currently has an importance that would have been hard to envision even a few years ago. The ferment about educational performance that began more than a decade ago continues to grow, and each wave of reform proposals further increases the importance accorded to testing. These trends are now expressed at the national level both in proposals for the NAEP and in proposals outside the federal government for the development of additional national tests that would be used--unlike the NAEP--to test all students.

To judge how well the NAEP and other tests can meet the rising expectations policymakers hold for them, it is necessary to consider three basic issues: the *multiple functions* that the tests are being asked to serve, the problem of *corruptibility*, and the *unavoidable technical complexity* of large-scale assessments.

Currently, tests are being asked to serve many purposes, three of which are central to proposals for large-scale assessments such as the NAEP. These are:

- o *Monitoring:* providing a description of achievement.
- o *Evaluating* educational programs or policies: not just describing *what* students know, but also determining *why* they know it.
- o *Holding people* (students, teachers, administrators, or those holding political office) *accountable* for achievement.

Monitoring is the traditional function of the NAEP; it was designed to inform debate and improve education by providing the nation's best description of the achievement of its youth. Traditionally, the NAEP had no *evaluative* or *accountability* functions, but recent changes and proposals, such as the movement toward state comparisons, rest in large part on proponents' desires for a national test that can be used to evaluate programs and policies and hold to hold people accountable for them.

Unfortunately, it is very difficult for one test, no matter how good, to serve all three functions simultaneously. First, the different functions call for different designs. For example, for purposes of monitoring, it is acceptable to sample students at a single point in their schooling and to use resources primarily to insure adequately broad sampling of both students and tasks. To evaluate programs, however, it is necessary to channel resources differently, to obtain more detailed data on factors that influence test scores and even to track the growth of individual students over time. Second, using a test for accountability can undermine its accuracy, thus undermining its usefulness for monitoring.

This points to the second issue: the *corruptibility* of tests. Many people incorrectly view tests as an incorruptible measure, like a fever thermometer. If a fever thermometer shows a drop in temperature, one can be pretty sure that the fever itself has declined. Many other common measures, however, are more easily undermined. On example might be airline on-time statistics. When an airline's on-time statistics improve, you might well ask why. Has the airline in fact decreased delays, or has it simply increased the estimated time of its flights, to allow for long delays on the tarmac while remaining "on time?" If the latter, nothing has really happened to how long it takes to get to Detroit.

Similarly, tests can be corrupted; particularly when they are used for accountability, *scores* can increase more than actual *achievement*. This is one basis for the now well-known "Lake Wobegon" phenomenon: the fact that most states and districts now report themselves to be "above average." One reason this can happen is simple: *teaching to the test*. Tests are necessarily a very small sample of what we want students to know. We can't test their knowledge of tens of thousands of words, so we test them on a sample of 40 or 60. As long as the 60 are a reasonable sample and no one directs teaching at those particular words, this works: one can infer the level of mastery of thousands of words from mastery of the 60. What if, however, a teacher simply has students memorize the 60 words? Scores skyrocket, but total vocabulary is scarcely affected. This is a particularly extreme example, but less extreme teaching to the test can also inflate scores on the test significantly.

The final general point that must be borne in mind is that large-scale assessments such as NAEP--if they are to provide us with useful, valid, and reliable information--are necessarily highly complex. Designing and implementing them requires substantial technical expertise in measurement and sampling. Even issues that seem on their face to be largely substantive rather than technical can have complex technical aspects that must be addressed if the test is to provide meaningful information. Currently, there is widespread impatience with the technical aspects of measurement in much of the education policy community, but that impatience does not lessen the impact that technical considerations have on the information we obtain.

With these basic points in mind, I would like to discuss the operation and statutory basis of the NAEP.

MONITORING

The NAEP remains ideally suited to its traditional role of monitoring the achievement of the nation's youth. Moreover, its value for this purpose is hard to overstate, because nothing else is presently a close substitute for it. We have no other assessment that tests nationally representative samples of students at frequent intervals on a broad range of subject matter that we believe important. In addition, unlike many state-level tests, the NAEP is not *yet* taught to in any way that would undermine its results.

Nonetheless, the NAEP could be improved to better serve this monitoring function. Some of the potential improvements would require resources that are currently very scarce, both because of the general budgetary situation and because the NAEP's resources are being siphoned into other uses, such as the futile effort to evaluate states' educational programs.

The improvements that might be made to the NAEP to monitor achievement better are many, but these are a few possibilities:

- Adding periodic supplements to assess high-achieving students, especially in mathematics and science. This group is a major focus of policy debate, but the NAEP, designed to assess the student population as a whole, lacks both the appropriate sample of students and the appropriate mix of tasks to assess this group well.

- Augmenting the sampling of other groups of importance to policymakers, such as immigrant students and college-bound minority youth.

- Strengthening the currently very limited array of background information collected to describe students and schools.

- Further expanding of the range of knowledge and tasks represented on the test.

The last of these possible improvements points to a serious deficiency in the current administration of the NAEP: the insufficient weight given to important technical considerations in making certain decisions about the operation of the NAEP. The NAEP is moving rapidly to incorporate new forms of assessment (particularly, various types of "performance assessment") and to decrease its reliance on multiple-choice questions. While this effort is laudable, it is difficult; we don't know as much about alternative forms of assessment, and the NAEP has to invent things as it goes along. The difficulties inherent in making this change have not been taken into account adequately in charting the NAEP's operation.

Currently, the Educational Testing Service (ETS), which is the prime contractor for the NAEP, is neither expected to evaluate these new forms of assessment adequately nor provided resources to do so. Rather, ETS is pressured to incorporate new assessment methods into the NAEP in the shortest possible time, even in the absence of any confirmation that they are of sufficiently high quality to be informative. At a recent meeting of a NAEP technical advisory committee, Archie LaPointe, director of the NAEP, drew the accurate analogy of trying to retool the engine of a car while it is moving at 60 miles per hour.

This specific problem could be addressed by changes in the authorizing statute. For example, the statute could provide for a period of time--three years was a recent suggestion by some technical advisors to ETS--for new assessment methods to be developed and evaluated before they are incorporated into the NAEP for purposes of national reporting. Similarly, the statute could direct NCES to provide a reasonable set-aside of NAEP funds to be used for this purpose. (While there is room for debate on this point, it might be most efficient to divide these funds between the NAEP prime contractor and other independent research groups.) These changes would slow the adoption of new forms of assessment but would assure that those finally used are worthwhile. The more general problem of insufficient consideration of technical issues can also be addressed by statutory changes; I will discuss those shortly.

Ironically, an additional way to improve the monitoring function of the NAEP would be to increase funding for competing assessments. There are two reasons for this. One is the competition *per se*, which might over time lead to more effective or efficient ways of conducting the NAEP. The second reason, however, is the one I want to stress: fielding additional assessments will enable us to interpret the results of the NAEP with more clarity and certainty. Because tests are only samples of what we want to assess and are technically complex, even fundamental results can differ substantially from one test to another. Put simply, it is risky to put too many eggs into any one testing basket, and even the NAEP can occasionally mislead us. The trustworthiness of the information provided by the NAEP therefore could be bolstered by increasing the amount of information from alternative sources. Clearly, additional assessments of the scope and frequency of the NAEP are not practical, but they are not needed for the purpose of validating the NAEP. The information needed might be obtained, for example, by funding periodic special studies and by augmenting the testing associated with the Education Department's periodic longitudinal surveys.

EVALUATION

The current controversy about using NAEP for state comparisons hinges in large part on the reasonableness of using NAEP to evaluate policies and programs. State-level *monitoring*--providing descriptive information about the severity of educational problems--could be useful, but in fact few proponents of NAEP state comparisons appear to want that. What they want is information about the relative effectiveness of states' educational programs.

The NAEP, however, as it is currently designed, *cannot provide reliable information about the relative effectiveness of state programs.* It can tell us which states have higher achievement, but not *why* they do. In some cases, State A will outscore State B because of better educational policies, but in other cases, the cause will be something else entirely, and we will rarely know which is which. In addition, when State A does outscore State B for educational reasons, we will often not know *which* educational factors are the cause. Precisely the same uncertainties would arise at the local level if the current statutory prohibition against using NAEP to evaluate or compare schools and districts were lifted.

The reasons for this are too complex to discuss in detail here, but they boil down to one essential point: the NAEP does not provide the information we would need to rule out reasonable alternative explanations of differences in scores or to pinpoint educational factors that might be responsible. Because the resources of the NAEP are channeled into providing the best affordable estimate of achievement, they are *not* channeled into producing other information needed for evaluation, such as detailed information on family and community background factors, educational policies and practices, and students' prior levels of achievement.

I have been in debates in which the proponents of NAEP state comparisons insist that the NAEP is adequate for drawing inferences about educational effectiveness, but in making that argument, they are accepting a standard of evidence far weaker than we routinely require in areas outside of education. The methods for evaluating educational programs are basically the same as those used to evaluate, for example, medical treatments or health programs. In those fields, data similar to the NAEP would not be considered adequate for reaching conclusions about effectiveness or safety. If we accept weaker evidence in the area of education, we may be less likely to do someone bodily harm, but we are just as likely to be wrong. Just as looser standards of evidence would sometimes cause medical practitioners to continue ineffective treatments or to terminate effective ones, such standards would likely mean that some ineffective educational programs will gain credit and be emulated, while some effective ones will be threatened.

The NAEP could be redesigned to support evaluation of educational programs, but it would require wholesale changes in the program. Moreover, the changes that would be called for--such as obtaining detailed information on family background, school programs, and, especially, on students' achievement growth over time--are likely to be very expensive. Absent significant new resources, NAEP could not assume those additional roles without substantially weakening its collection of the basic achievement data that it needs for purposes of monitoring.

ACCOUNTABILITY

The appropriateness of using the NAEP for accountability also arises in the debate about state comparisons and is a key consideration in the debate about the current statutory prohibition against the use of the NAEP for comparisons of local districts or schools.

To some extent, using NAEP for accountability purposes would run afoul of the same difficulties just noted, for it is hard to hold people meaningfully accountable for differences in educational achievement if you cannot show that they are in fact responsible for them. There is, however, another reason to avoid using NAEP for accountability: it could undermine its value for monitoring.

The key issue, as I suggested earlier, is teaching to the test. Whether, and under what circumstances, teaching to the test is instructionally desirable is the subject of an intense debate in the education community. For present purposes, however, you need not be distracted by that debate.

Whatever the instructional danger--or value--of teaching to the test, it can seriously undermine the validity of test scores as an index of achievement.

Some proponents of using NAEP for accountability maintain that NAEP incorporates protections against this kind of corruption. It does. As currently operated, NAEP provides considerably more test security than do many other testing programs. In addition, it uses a sampling system in which each tested student is given only a fraction of the test. Both make it more difficult for teachers and students to learn what the test contains, thus discouraging inappropriate teaching to the test. It is by no means clear, however, that these protections are adequate. Even though NAEP tests are collected after use, many thousands of students and a good many teachers see the test while it is administered, and others see it for other reasons. Over time, knowledge of the content of the test will grow, and at that point the corruption caused by teaching to the test can begin.

Local comparisons are far riskier than state comparisons in terms of potential corruption of the NAEP. In all but the smallest states, NAEP state comparisons leave most students and schools untested. Thus the people who are responsible for much teaching to the test--local educators--will often have no direct incentive to raise scores. State administrators will, but they have fewer means of pressuring local educators than in a system where the students of local educators are tested. Local use of the NAEP would dramatically increase the number of people whose performance (that is, whose students' performance) will be measured and will thus greatly increase both the incentive to raise scores and the means of exerting pressure to do so.

Exactly so, respond the proponents of local use of NAEP: it would give people more incentive to improve their performance. It indeed could, but at the cost of threatening the fundamental validity of the descriptive information that NAEP was created to supply.

This dilemma has a solution: use the NAEP for monitoring, and use other tests for accountability. If proponents of local comparisons--and the audiences who would get the results of them--were convinced of the necessity of separating the accountability and monitoring functions of tests, it might be possible to have local reporting of the NAEP while not using it for accountability purposes. Under current conditions, however, that prospect is unrealistic. Therefore, the prohibition against local use of NAEP results is still needed to maintain the NAEP as an instrument of monitoring rather than accountability.

THE ROLE OF THE NATIONAL ASSESSMENT GOVERNING BOARD

I would like to make only a few comments about the National Assessment Governing Board (NAGB) in the context of the third of my general points: the unavoidable technical complexity of the NAEP. How should the NAEP be administered to insure that the technical ramifications of issues such as state comparisons and the use of performance assessments are acknowledged and handled appropriately?

Under the current statute, NAGB has extremely broad authority over the design and conduct of the NAEP and over the reporting of NAEP results. This authority includes issues that can be resolved without a great deal of specialized expertise, such as selecting the subject areas to be assessed. Indeed, some people have argued--reasonably, in my view--that decisions of this sort should *not* be considered the province of experts; although experts should be consulted, these issues are appropriate ones for wide public debate. On the other hand, the authority of NAGB also includes issues that are technically very complex, such as "developing test specifications" and "designing the methodology of the assessment."

Under current law, NAGB is not constituted appropriately to deal with technical issues of this sort. Only two of the 23 seats are set aside for experts in testing and measurement. NAGB has also not used the leeway it does have under current law to augment its own limited technical expertise. For example, it has not constituted a standing panel of technical experts to advise it. (By contrast, ETS, as prime contractor for the NAEP, does have such a panel, despite the impressive expertise of many of its own staff.) NAGB does make use of outside technical consultants, but only on an *ad hoc* basis. These *ad hoc* arrangements do not provide a regular opportunity for experts to meet and deliberate about major issues confronting the NAEP.

The contrast between NAGB's level of technical expertise and that of other government-funded groups with authority over technical decisions--for example, NIH or ADAMHA review panels--is striking. We would not assemble a panel to design clinical drug trials that included only a few people with technical expertise in designing clinical trials, yet we have done precisely that in constituting NAGB.

This lack of expertise is extremely important for several reasons. First, in educational measurement, as in other complex technical fields, there is no clear way to separate technical and substantive questions. In pharmacology, "merely technical" considerations determine whether the medication you take will cure you or harm you; in measurement, "merely technical" considerations determine what the results of a test mean. As presently constituted, NAGB does not have the ability to resolve some important technical questions. More important, in some instances, it may not have the expertise needed even to judge when technical issues are in fact important. To pick only one example, the organization charged with overseeing the NAEP should be able to recognize the need for validation of new assessment methods before they are fielded and reported on a national basis.

Second, NAGB needs technical expertise in order to be able to respond to the increasingly extreme political pressures being exerted on assessments by various constituencies. These pressures are often very intense, and the constituencies involved appear to have ever less patience for technical issues that might slow changes they desire. Expertise in measurement is sometimes needed to understand which political pressures are reasonable and to find technically sound ways of accommodating them.

The quality and effectiveness of the NAEP therefore could be improved by changing the statute to insure a reasonable representation of technical expertise on the National Assessment Governing Board. Certainly, NAGB needs to incorporate perspectives of other sorts, because the goals of education are properly a matter of public debate, not the sole province of experts. The design of assessments, however, does require specialized expertise.

The extent to which the membership of NAGB should be modified depends on its purview. If NAGB retains authority over the highly technical issues it now oversees, its membership should be altered fundamentally. If, on the other hand, authority over some technical issues were removed from NAGB's purview, more modest changes might suffice. Even then, however, NAGB would need sufficient technical expertise to be able to judge the importance of technical issues and to appreciate the technical ramifications of initiatives it considers.

Among the steps that might be taken to strengthen NAGB are these:

- o Increasing the number of seats (of the 23 total) assigned to technical experts in testing and measurement to provide an effective voice for technical concerns. With NAGB's current purview, a reasonable level might be roughly a third of the Board--say, 8 of 23 seats. If NAGB's authority over technical decisions were reduced, it might suffice to allocate fewer seats to technical experts--perhaps 4 or 5.

- Requiring that the individuals chosen for those seats be widely acknowledged as experts by their professional peers. There is currently no such requirement in the law. NIH and ADAMHA review panels might provide a useful example of how such expertise could be assured.

- Requiring that NAGB constitute a standing advisory committee of experts in testing and measurement. This committee should meet several times a year and should be empowered to issue reports to NAGB, NCES, and other interested parties to the extent the committee deems necessary. This committee could presumably overlap with the technical advisory committee set up to advise the NAEP prime contractor.

CONCLUSION

The NAEP is currently at a turning point. With the currently authorized state comparisons, NAEP is taking major steps toward the functions of evaluation and accountability. Proposals for further changes in these directions have been on the table since the last Congress, and as the reauthorization of NAEP nears, pressures for further changes will intensify.

My own view is that the changes made to date were undesirable and that further changes in those directions--such as the use of NAEP for local evaluation and comparisons--should be avoided. Absent dramatic changes in the structure of the NAEP, the efforts to use it for evaluation are futile and potentially misleading, and efforts to use it for accountability threaten the NAEP's traditional role as the nation's only frequent and nationally representative indicator of the achievement of our youth.

The Committee will hear--indeed, has already heard--fundamentally different views of what NAEP should be, and you may well decide that the traditional monitoring function is not in your view the most important role the NAEP can play. I would ask you to bear in mind, however, that despite some contrary claims, we cannot have our cake and eat it too, and pursuing evaluation and accountability will at best require short-changing the monitoring function and at worst will undermine it.

Finally, I would like to reiterate that providing the information on achievement that policymakers and the public wants is a technically complex and difficult task. If the NAEP is to be successful as it evolves further, it must be given a governance structure that is cognizant of these difficulties and able to balance them against legitimate political pressures.

Mr. Chairman, thank you again for the opportunity to discuss these issues with the Subcommittee. At this time, I would be pleased to turn to any questions that you or other members of the Subcommittee have.

Chairman KILDEE. Thank you, Dr. Koretz. Dr. Linn?

Dr. LINN. Thank you, Mr. Chairman, for inviting me to testify today. I, too, will try to be brief and ask that my written comments be admitted into the record.

Chairman KILDEE. Without objection, they will be included.

Dr. LINN. Many of the points that I want to make have already been made, and I would like to just reemphasize a few of those.

To begin with, NAEP was designed, as has been pointed out several times, not as a test in the usual sense that you give to get scores on individual children or on individual schools but as an information system to monitor progress of the Nation's youth in educational achievement over time.

It has served us quite well over the last 20 years. We haven't always liked the messages that it has provided such as achievement being—the trends being rather flat and at levels that are below those that we aspire to as a nation, but nonetheless it is the best source of information that we have to have an independent monitoring of what is happening to achievement in the Nation.

The current legislation has prohibitions against reporting scores for individual students or for individual schools. I would strongly urge that those prohibitions remain in effect. I think it is just impossible for the system to serve both as a monitor of trends of achievement for the Nation and as an instrument that will report scores for individual students or for individual schools.

Currently, the sampling of exercises for an individual student is quite slim, too slim to provide reliable scores for an individual student. To have reliable scores would greatly increase the amount of testing that would be required as well as distort the function of the monitor as an independent monitor.

The same could be said for individual schools. By sampling and giving different exercises to different students, you can get a good reading for the Nation as a whole or for States as in the case that you now have permitted so I would strongly urge again that that prohibition against going down to the student or school level remain in effect.

One of the major changes that took place in the 1988 legislation, of course, was the provision to allow the voluntary participation of States in the national assessment. In 1990 in accord with that for the first time, 37 States did participate in that in eighth grade mathematics, and in 1992 a large number of States plan to participate in both reading and math at grade four and mathematics again at grade eight.

Congress in authorizing the trial State assessment put in place a requirement that there be an evaluation conducted. That is currently under way under the National Academy of Education's guidance. I have the pleasure of co-chairing that panel, and that panel will be reporting in October as mandated by Congress.

Because Congress needs to consider the authorization at this point in time, the panel will be submitting an interim report in about three weeks. I cannot at this point give you the details of that report because it is still under review, but I can say a couple of things.

First, we have not identified major flaws in the administration of the National Assessment at the State level that would preclude the release of the scores in June.

Second, we have not seen adverse effects on the National Assessment itself, and so will be recommending that some kind of trial continue in 1994. Why continue a trial in 1994 as opposed to a full operational program? In part because many of the—while you will have tested in two grades in mathematics and one in reading, there are still many things that will be untried at the end of 1992, much less at this point, in 1990.

In particular, there will have been no trial with twelfth grade students, and it is likely that just because you are successful at eighth grade or fourth grade, it does not necessarily follow that you would be successful at twelfth. The motivation to participate on the part of students may be quite different at twelfth grade than at other grades.

Moreover, the success on trials at a couple of grades in only two subjects in one, and one subject in another, is far removed from a full operational program that would involve three grades and three or four different subject areas, so we think that there are still a number of issues to be considered.

In addition, there are parts of the trial State assessment that is limited to only public school students at the current time, and if the goal is to really have comparisons of performance of all students and States or States to the Nation, the number of students in private schools varies considerably from State to State, and that would be something that the Congress might want to consider.

Finally, returning to national NAEP, I would like to point out one of the things that the original designers of National Assessment had in mind, and that was the assessment of all of the Nation's youth at a given age level.

Early on, that included students in public schools, in private schools, and youth who had dropped out of school. Due to budget considerations and constant dollars, the out of school youth are no longer a part of the National Assessment. This is particularly bothersome at age 17 where, in fact, we are assessing only those students who are in schools and missing about a quarter of the 17 year old youth that are no longer in school.

So I would urge consideration of the need to know how not only students in school are doing but 17 year old youths in general.

I was asked to comment on the National Assessment Governing Board. I claim no expertise in government so I will not make any recommendations there.

I will make one observation, though, and that is that there seems to be some ambiguity in the degree of responsibility and authority between the National Assessment Governing Board and the Secretary of Education. I think that some clarification of that would be desirable.

Thank you, Mr. Chairman.

[The prepared statement of Robert L. Linn follows:]

Statement of

Robert L. Linn
Professor, University of Colorado at Boulder

before the

Subcommittee on Elementary, Secondary, and Vocational Education
Committee on Education and Labor
U.S. House of Representatives

March 13, 1991

Mr. Chairman, thank you for the opportunity to provide testimony to the Subcommittee oversight hearings regarding the National Assessment of Educational Progress (NAEP). The hearings and consideration of legislation regarding NAEP have particular importance because of the current national attention that is being given to educational goals and the role of assessment both as a means of monitoring educational progress and as an instrument of change.

The inclusion of the word "progress" and the choice of the word "assessment" rather than "test" in naming NAEP were not accidental. When the original plans were being developed for NAEP roughly a quarter of a century ago by people such as Ralph Tyler and the late Francis Keppel, NAEP was conceived of, not as a test in the familiar sense, but as an information system, one that would yield indicators of educational progress. The Gross National Product and the Consumer Price Index were used as analogies to the type of indicator that was needed to monitor progress in the educational achievement of the nation's youth.

The role of NAEP as an independent indicator of educational progress is quite different than other roles, such as school and teacher accountability or the certification of individual students, that tests are expected to serve. NAEP has provided the most reliable single source of information about the trends in achievement of the nation's youth over the more than twenty years since the administration of the first assessment. Although we may not like some of the messages that NAEP has provided (e.g. levels of achievement that have changed relatively little during the past twenty years and remain below those to which we as a nation aspire), we have no better way of monitoring educational progress for the nation as a whole.

There are, of course, other roles for assessment. This is apparent in the current debate regarding the possibility of a national examination system or a national test. It is, however, critical that the different purposes be clearly distinguished and that the assessment systems be designed to be consistent with those purposes.

Prohibition Against Ranking, Comparing, or Evaluating Individual Students, Schools, or School Districts

NAEP was not designed to be a test that would provide scores for individual students or schools. Indeed, such uses are precluded in the current law which requires confidentiality of individual student identification and of results for individual schools. Ranking, comparing, or evaluating, individual students, schools, or school districts is also prohibited.

I strongly urge that individual student and individual school results continue to be kept confidential in future legislation. An expansion of NAEP to provide results for individual schools or students would lead to a loss of NAEP's value as an independent and uncorrupted indicator of educational progress.

Since each student takes only a sample of the exercises used at any given administration, NAEP does not collect enough data to produce reliable scores for individual students. The administration of different samples of exercises to different samples of students is a strength for purposes of obtaining a broad indicator of performance for the nation or a state, but this use of multiple samples of exercises would be a disadvantage for producing scores for individual students.

The reporting of results for individual schools would greatly increase the size of the sample that would be required at each participating school. More importantly, the focus on school building results would increase the stakes that are attached to NAEP results, degrade the security of NAEP exercises, and increase the likelihood that results would be corrupted. Such a change would threaten the integrity of NAEP as an independent monitor of national or state trends in student achievement.

The extension of NAEP beyond the state level to the local district level raises a somewhat different set of issues. Although I do not advocate an expansion of NAEP that would allow voluntary participation by large school districts, I believe that the case for allowing participation may be as strong for large school districts as it is for individual states. The current design which requires the assessment of only a sample consisting of a relatively small fraction of the students within a grade applies as readily to large districts such as Los Angeles, Houston, or New York City as it does for a number of states. Indeed, a number of districts would have more students than some of the less populous states. Because of its special status, one district, the District of Columbia, did participate in the 1990 Trial State Assessment. Consequently, it seems just as reasonable and feasible, at least on a trial basis, to allow large districts to participate voluntarily in NAEP and receive district level reports as it is to provide this option to states.

If the prohibition against the voluntary participation of school districts is removed, I believe it would be wise to do so on a trial basis during which time the practice could be evaluated much in the way that this is now being done for the Trial State Assessment. I also believe that it would be wise to maintain a restriction for districts with enrollments below some pre-specified level.

Trial State Assessment

One of the most significant changes in NAEP introduced in 1988, when Congress enacted Public Law 100-297, was the introduction of the Trial State Assessment (TSA) program. Consistent with the provision of the law, NAEP made available to the states an eighth grade mathematics assessment in 1990. Thirty-seven states, the District of Columbia, and two territories voluntarily participated in the 1990 TSA. Results of that first TSA are scheduled for release in June of this year. Planning for the second TSA, which will be expanded to include both reading and mathematics at grade four as well as mathematics at grade eight, is well underway.

In accordance with the mandate of the law, the National Center for Educational Statistics awarded a five-year grant to the National Academy of Education to conduct an independent evaluation of the Trial State Assessment. That evaluation is being conducted by an eighteen member panel with a wide range of technical and policy expertise that was appointed by the National Academy of Education. I have the pleasure of serving as co-chair of the panel with Professor Robert Glaser from the University of Pittsburgh.

The panel is scheduled to report on the evaluation of the 1990 Trial State Assessment in October, 1991. A second report, which will include an evaluation of both the 1990 and 1992 Trial State Assessments, is scheduled for late fall of 1993. The panel has commissioned a number of analyses and papers that will, as the law requires, "assess the feasibility and validity of [state] assessments and the fairness and accuracy of the data they produce."

Although a complete evaluation of the Trial State Assessment at this time--almost three months before the first state results are scheduled for release--is not possible, it is recognized that Congress needs to consider authorizations now if state assessments are to be continued in 1994. Hence, the National Academy of Education Panel will submit an interim report within the next few weeks.

Because the interim report is still under review, I cannot provide details of that report at this time. However, I can say that the panel has not identified major flaws in the sampling or administration of the Trial State Assessment that should preclude the release of the state results as scheduled in June. I can also report that the initial experience was sufficiently positive and without adverse effects on the national assessment that the panel will recommend continuation of an expanded trial in 1994.

Even at the end of the 1992 assessment there will be several aspects of NAEP that have not been evaluated for use at the state level. Only fourth and eighth grade students will have been included in the TSA with assessments in only one subject area at grade eight and two at grade four. Only public school students will have been assessed in the TSA.

There are significant differences between high school seniors and students in grades four or eight in their likely willingness to participate in an assessment, their degree of cooperation during the assessment, and their motivation. These differences make it unwise to assume that a successful trial at earlier grades implies success for a twelfth grade state-by-state assessment. Similarly, successful implementation of state assessments in one or two subjects at two grades does not necessarily imply that a full-scale, operational state-by-state assessment in three grades and three or four subjects, could be conducted successfully without degrading the quality of the national assessment results. Finally, success with only public school students does not guarantee success if both private and public school students are included in state level assessments. Hence, it seems prudent to authorize a continuation of an expanded trial through 1994 rather than a full-scale implementation.

Because there is considerable variability from state to state in the proportion of students enrolled in private schools, exclusion of private school students from the state samples degrades the comparability of state-to-state results. A better indication of the achievement of the students within a state would be obtained if both private and public school students were included in the state administrations of NAEP. Therefore, I believe that provisions for the addition of private schools to the state samples deserves consideration.

Out of School Youth

In the early years of NAEP, assessments were administered to age cohort samples, both in and out of school. Due to declining support in constant dollars, the inclusion of out-of-school 9, 13, and 17-year-olds was lost. This loss is particularly serious for the 17-year-old cohort.

It is important to know what high school seniors can do. It is also important to monitor the achievement of 17-year-old students who remain in school regardless of grade level. Even the latter, however, excludes roughly 25% of the nation's 17-year-olds who have dropped out of school.

The criticality of including out-of-school youth in the assessment was stressed in the original design of NAEP. The rationale was clearly articulated by Ralph Tyler in 1966. "In our total society, there are children in public schools, in private schools, and some who are not in school. Hence, a comprehensive assessment program requires samples which include children who are found in these several places."

I believe that Tyler's reasons for wanting out-of-school youth included in the assessments are, if anything, more compelling today than they were a quarter of a century ago. The use of college admissions tests as indicators of educational progress is much maligned, and rightly so, in part, because these tests are taken by only a selected and changing segment of the population that is planning to go to college. Although less severe, NAEP is subject to a similar criticism, especially for the 17-year-old cohort.

The reintroduction of the assessment of out-of-school youth as a regular feature of NAEP would provide for a substantial improvement of the quality of information that NAEP provides about the educational progress of the nation. For purposes of providing a national indicator of achievement, a kind of Gross Educational Product, I believe that the inclusion of out-of-school youth deserves high priority.

National Assessment Governing Board

One of the topics mentioned in Congressman Kildee's letter of invitation to testify was the role of the National Assessment Governing Board (NAGB). Since I claim no special expertise on the subject of governance or bureaucratic organization, I will not make specific recommendations regarding the role of NAGB. I will make one observation, however.

There appears to be considerable ambiguity in the extent of NAGB's responsibilities and authority vs. those of the Secretary of Education. Clarification by Congress seems desirable, and once clarified, it will be important that funding is commensurate with the designated responsibility and authority.

Respectfully submitted by Robert L. Linn for the March 13, 1991 hearing of the Subcommittee on Elementary, Secondary, and Vocational Education of the Committee on Education and Labor, U. S. House of Representatives.

Chairman **Kildee**. Thank you very much.

Dr. Roeber?

Dr. **Roeber**. Thank you, Mr. Chairman. I, too, request that my written statement be entered into the record.

Chairman **Kildee**. Without objection, it will be entered in the record.

Dr. **Roeber**. I also apologize in advance if my voice fades. I came down with, I think, a case of the flu coming back from the West Coast at a meeting last week, and sit here in a rather weakened condition. I wish I was stronger so I could speak out more forcibly on this topic.

As a director of assessment in Michigan since the mid 1970s, I have had the opportunity of working with States in their programs and helping to shape State programs since the mid 1970s, but since I began my career working for National Assessment in the early 1970s, I have seen National Assessment from a personal perspective and can bring that to your discussions.

I support the current shape of National Assessment as a monitor of national achievement in our country, and I feel strongly that that is an appropriate role to continue for National Assessment.

I, too, have concerns about the ways in which a State NAEP, a district NAEP, a school NAEP, or a student NAEP, might unfold. These deal with the substance of National Assessment as well as the process for arriving at that.

First, the process concerns. I am concerned that national programs such as the National Assessment or the others that you will consider tomorrow are being developed without the involvement of State level curriculum and assessment persons. We have had an assessment program in Michigan for 22 years. Most of the lessons that are now being learned and/or rediscovered at the national level we have learned the hard way in Michigan, and that is one of the reasons in the mid 1970s we formed the Association of State Assessment Programs so that we could share those hard knocks and prevent others from going through that.

I think the notable exception to this has been the consensus process that the Council of State School Officers has been using to arrive at the assessment frameworks.

Most specifically, we have objected to the ways in which the National Assessment Governing Board has adopted plans for the future which call for sweeping changes in the scope of National Assessment without State involvement, and certainly the technical expertise that was referred to earlier is something that is desperately missing from that so that I have felt strongly for a couple of years that practitioners such as State curriculum people, State testing people, as well as our counterparts at the local level, need to be heavily involved in the design of the National Assessment or other National Assessment programs.

We have been involved to some extent in the administration of the State National Assessment program, but we seem to be most involved in the trivial aspects of the administration and almost completely uninvolved in the design, in the shaping of the policies that underlie it, as well as the procedures for assessment. I think we have much to offer.

Now my substantive comments about the future of National Assessment. First of all, I am concerned about the quality of the assessment materials that are used. There is a trend now nationally to incorporate performance assessments in assessment, and I urge the National Assessment Project to go back to the future.

When I worked in National Assessment in the early 1970s we had individual performance testing in areas as diverse as art, music, science, social studies, career and occupational development, mathematics, and writing. We literally assessed 40,000 or 50,000 youngsters in singing, playing musical instruments, and ended up with about 80,000 tapes which we scored, and so the National Assessment Governing Board's comments that relate to the use of performance assessment, that it is untried, and that we need to go slow on this, to me are simply not a recognition of the heritage of the project that they are supervising.

Second, I think the most valuable aspect of National Assessment, it is an original design. I think a real testimony to Ralph Tyler's inspiration in designing this program in the mid 1960s was the inclusion of an assessment of young adults, age 26 to 35.

Almost no one knows about that because it was dropped in the early 1970s because of Federal budget concerns. The reason why I feel that this is the most valuable part of National Assessment is because what ultimately we are concerned about in this country, and rightfully so, is whether our high school graduates, our college graduates, and even those with advanced training, are prepared to enter the workplace and be productive citizens.

What better way to determine that than, in fact, to carry out an assessment of these individuals?

I think that, once again, it demonstrated the inspiration of the early design of National Assessment that this 20,000 to 25,000 student or adults in that age range could be tested on an annual basis in an appropriation, incidentally, that is about comparable to what it is now, unadjusted for inflation; that is, the appropriation in the early 1970s was about $6 million to $7 million a year, and they were able to do this individual performance testing that I referred to as well as that of the young adults.

Third, I am concerned about the content areas that are assessed. There seems to be an over-emphasis in my mind on mathematics, reading, writing, social studies, and citizenship, and I would like to see the National Assessment branch out into some other areas that I think are vitally needed, including the arts, foreign language, and health and physical education. I am sure that there are other areas.

Fourth, as I look at assessment above the level of the State, I am most concerned about whether and how Michigan performs relative to the world. You know, the automobile industry is not concerned whether the autos built in Indiana or Ohio are better than the ones built in Michigan. We are concerned about building better autos than the rest of the world. The same ought to be true with student achievement.

So rather than driving NAEP downwards into our schools and districts and even the State, I am concerned about driving NAEP upwards and linking to international assessments. We had the opportunity. Unfortunately, we couldn't come up with the funding. It would have cost the State $120,000 to link the Michigan reading test at fourth grade to the International Education Achievement reading literacy test at fourth grade, and I think that would have provided data for every school and every district on the IEA Reading Literacy Scale for $120,000 without having had to give that test to all of our students.

There are ways that we ought to be exploring to link the assessments without using one to drive out another, to link them together so that one assessment is essentially carrying the load of two or three, and, in fact, given what I could have gotten out of the IEA Reading Literacy Study, that is school, district, State, national, and international data, I really wondered whether we needed the National Assessment State by State data at grade four.

School and district testing on NAEP? It is certainly a controversial issue, and there are States that are in favor of it. There are States that use commercial available standardized achievement tests, and they would prefer to use the National Assessment. Other States, such as Michigan, find less value in that, and I understand that that is an issue that I think you are going to be wrestling with.

But I think that I personally am in favor of lifting that prohibition so long as all of the principal parties at the State and local level are in agreement to it so that it is not forced on someone. It is left on a voluntary basis at the State and the local level so that if the local district cares to get that information, they can, and if they choose not to, they also can avoid that.

Frequency of testing is an issue that you are going to be also discussing. I agree that annual assessments would be desirable, but I disagree about a two year cycle. NAEP used to be on a four year cycle, and I think that is frequent enough to see the kind of changes that we expect on a global basis in achievement in students.

Programs such as National Assessment have been quite able over cycles as long as five or six years to track trends in student achievement because, again, as a monitoring program testing every four years or five years is appropriate.

If the purpose of NAEP is an accountability program to force schools to change, then clearly four years is way too long, but I believe that NAEP's greatest value is a monitor.

The particular issue of the motivation of twelfth graders has been alluded to, and it is one that urges me to urge you to continue the trial nature of National Assessment State assessment programs for 1994.

Twelfth grade was selected as the third of the three grades to be assessed. Testing takes place in February. Anyone familiar with schools and how they operate will indicate to you that seniors in February are not very motivated.

Now, I am not particularly concerned about whether Michigan students are more or less motivated than Indiana students so in a State comparative program, I don't care. But when people are making absolute judgments about what our high school graduates have at the time they leave school, then it is a critical issue.

When National Assessment was assessing 17 year olds, those assessments, the modal grade level at which those assessments occurred was eleventh grade. Now, that is not quite as neat as having a picture of students as they leave school, but right now I am very fearful that the twelfth grade data that we get is seriously underestimating what our students know because of the lack of motivation of students to take the test.

The final issue I want to talk about is whether testing can change student learning. Obviously, as a supervisor of a program that has been in business for 22 years, I believe it can, but will the test itself cause schools to change? We have a long history in Michigan of knowing that it won't; that in fact you need to engage classroom teachers in the process of change through local school improvement programs, through the provision of instructional assistants, model instructional programs, and even model curriculum.

It is a lengthy process, one that involves considerable technical assistance to be provided to each and every school that is participating.

What has concerned me about the 1990 and the 1992 trial programs is that we are proceeding as kind of a mixed model. NAEP is a monitor, but NAEP is also for school accountability, and when we get to the actual "what are we going to do with the results," there seems to be a hands-off approach at the national level, that it is not appropriate to develop suggested instructional strategies in mathematics or even to indicate to the public that a revamping of mathematics instruction is necessary.

Now, somehow it is okay to test students and beat States and schools up, but it is not okay to help them, and, in fact, there has been no funding provided for the assistance that States and schools are going to need to improve their instruction. I and others in the assessment and curriculum groups have taken it upon ourselves on a voluntary basis, sometimes at other agencies' expenses, to try and develop resources that States can use to interpret the National Assessment results.

In separate correspondence to the National Center for Education Statistics, I have recommended that at least five percent of every assessment be set aside for the kinds of technical assistance and improvement activity that I believe should occur after every assessment.

Mr. Chairman, I appreciate the opportunity of addressing these concerns to the committee, and I like the others would be happy to answer questions. Thanks.

[The prepared statement of Edward D. Roeber follows:]

DONALD L. BEMIS
Superintendent
of Public Instruction

STATE OF MICHIGAN
DEPARTMENT OF EDUCATION
P.O. Box 30008
Lansing, Michigan 48909

STATE BOARD OF EDUCATION
DOROTHY BEARDMORE
President
GUMECINDO SALAS
Vice President
BARBARA DUMOUCHELLE
Secretary
MARILYN F. LUNDY
Treasurer
CHERRY H. JACOBUS
NASBE Delegate
DICK DeVOS
BARBARA ROBERTS MASON
ANNETTA MILLER

GOV. JOHN M. ENGLER
Ex Officio

Testimony Before the House Subcommittee on Elementary, Secondary and Vocational Education

Edward D. Roeber
Supervisor, Assessment and Accreditation
Michigan Department of Education

March 13, 1991

Mr. Chairman, members of the committee, distinguished panelists and guests. I am Edward D. Roeber, Supervisor of the Assessment and Accreditation Programs for the Michigan Department of Education. I am also Co-Chair of the Association of State Assessment Programs, a technical assistance network among the individuals who work in state agencies responsible for conducting various state-level student assessment activities. The network has served as a communication tool among states to assist one another with the development and implementation of our state assessment programs. The network was established in 1976 because states were considering and implementing a variety of state-level programs and we felt that by working together we could help one another in implementing our programs in the best manner possible. The need for this technical assistance network has only increased in recent years, given the demands for new programs at the state level, the calls for reform in the nature of assessment, as well as the new assessment programs developed or being discussed nationally.

At the outset, I wish to express to the committee the group's appreciation for the opportunity I have been given to speak about the National Assessment of Educational Progress (NAEP) as the discussion the continuation and expansion of NAEP takes place. I want to first express the support of the group, myself included, for the type of assessment which is represented by NAEP. From a personal perspective, this support grows out of having started my career in assessment with three years of work on the National Assessment project when it was under the direction of the Education Commission of the States, as well as my fifteen years of work in Michigan. In addition, I am serving as a member of the Evaluation Panel which Congress established as part of the authorization of National Assessment Trial State Assessment programs to be conducted in 1990 and 1992. In all of this experience, I believe that there is a constructive role which a national assessment of student learning can play in the improvement of learning opportunities for American children and youth. I believe that NAEP has already played such a role in the past and can and should continue to do so in the future.

However, I have concerns with regard to the manner in which the National Assessment project is being planned and implemented, as well as regarding several specific issues. Attached to this testimony is a recent memorandum that I wrote detailing the concerns that I have about the **process** of implementing the 1990 and 1992 National Assessment Trial State Assessment Programs, as well as the planning for 1994 and beyond. As can be seen, I am most concerned that the state-level assessment persons, the ones with the wealth of educational, technical and practical experiences, are not really involved in helping to frame the issues about the future of NAEP, developing options for each issue, nor deciding the desirable option. These concerns relate primarily to the **process of involvement**, that is, that the persons with the practical experience are not utilized in the design of NAEP.

In addition, I have expressed concerns about the substantive design of the National Assessment program. These concerns center on several major areas:

Performance Assessment The use of authentic types of assessments in large-scale assessment programs has emerged as an area of real need in the last couple of years. This has caused developers of such programs, the National Assessment Governing Board (NAGB) included, to study how such assessments can be included in their programs. In the long range plan for NAEP which NAGB adopted in December, 1989 and reaffirmed in November, 1990, performance assessment is mentioned with a cautionary note urging further study and careful implementation through a series of special studies.

While I'm not opposed to careful implementation of new assessment strategies, it was NAEP that pioneered such assessments in the late 1960's and early 1970's. While none of this early work remains in NAEP, there were at one time individual and group performance assessments in areas such as reading, science, social studies, art, music, career and occupational development and so forth. Tens of thousands of students and young adults were tested with these. If any program could rightfully claim that such assessments were feasible and technically sound, it is NAEP. Therefore, it is time to "reinvent the past" and return such assessments to National Assessment; it is not necessary to be so cautious!

Assessments of Young Adults It is somewhat ironic that federal funds are being used to develop new studies of adult literacy, when one of the most valuable aspects of NAEP, the assessment of young adults, was dropped in the mid-1970's due to federal budget cuts. It is time to reinstitute this assessment and tie it directly to the in-school assessments of National Assessment. NAEP has already proven the feasibility and value of assessing the residuals of schooling among young adults. Such data would be invaluable in determining the readiness of students for college and careers. Rather than fund separate studies of adults through the Department of Labor, it seems to me more appropriate to carry these out in a fashion connected to NAEP.

Content Areas Assessed There are other important areas which should be included in the National Assessment of the future. These include: the arts, foreign languages, health and physical fitness. Each of these areas has been assessed in one or more states and several states have assessed them using individual and group performance exercises. In addition, NAEP has conducted two national assessments of both visual art and music, including the performance aspects of each program.

Motivation of 12th Graders One issue which has been raised repeatedly is the motivation of twelfth graders for the NAEP assessment when it is given in the spring of their senior year. In a comparative program, this issue is less important. After all, the seniors in one state probably won't be any more motivated than those in another. However, when the data is used in a more absolute sense to say what proportion of seniors appear to be well prepared for college or work, the lack of motivation on the assessment can cause serious under-estimates of their learning and could cause the nation to make serious errors in judgement about the changes which we must make to our educational systems in order to help students learn what they need for further education or work.

This issue deserves special study. The study is needed before further assessment work is carried out at grade twelve. In the past when students were assessed at age seventeen, most of these students were in grade eleven. Careful study of student motivation is needed before we invest so much in collecting grade twelve data that the longitudinal trends become so important that we can't wish change, regardless of how faulty the data may be. This is a serious challenge to the integrity of the data which needs careful study in the 1992 or 1994 NAEP. I have volunteered Michigan as a test site for this study, but haven't been able to convince the National Center for Educational Statistics (NCES) of the need for the study.

Assessment at the District and School Levels. One feature built into the 1990 and 1992 National Assessment Trial State Assessment programs is the prohibition on the release of the NAEP results by district, by building or by student. NAGB has proposed lifting this ban; several states are concerned about the impact of NAEP within their states and have asked that the ban not be lifted. Other states, wishing to replace their statewide norm-referenced testing program, have supported lifting the ban. Complicating this issue are the remarks made by the former chair of NAGB; these remarks clearly indicate the intent of NAGB to see that an every-pupil NAEP replace the variety of assessment programs which exists at the state and local levels.

It is my feeling that the ban should be lifted, but that Congress should do so in a manner which sets up a process by which this is carried out at the state level. This should be an voluntary activity at both the state and local levels, to be carried out at state or local cost, and with the approval of all policymakers at the level where the activity is to be carried out. This means the governor, state board of education and the legislature at the state level and the board of education and superintendent locally.

In addition, the NCES, NAGB and the NAEP contractors should be authorized to conduct statistical linking studies where states and local districts wish to engage in such work. This can serve the useful purpose of providing an opportunity for a state to report National Assessment scores without having to give NAEP tests to all students, nor having to link such tests annually. Where possible, the testing which takes place at one level should not drive out testing at another level. Instead, the efficiencies of linking such programs needs to be explored. It is simply inappropriate to assume that just because a test is administered nationally that it is better than those given at the state or local levels!

Frequency of Testing. NAGB has proposed shifting NAEP back to an annual assessment program, assessing six subject areas total or three areas per year on a two-year cycle. As long as NAEP is returning to the past in this regards, I strongly suggest that NAEP consider annual testing but on a three- or four-year cycle. This would allow a longer development cycle (three or four years instead of the current two years), more subject areas to be assessed (nine or more instead of six), and allow more time for trends to occur. It is doubtful that significant changes in student performance will occur in just two years. In the early years of NAEP, three- and four-year cycles were used to allow changes in student performance to occur before re-assessment. The assessment development cycle, currently at two years, just is not adequate for the types of development, pilot-testing, assessment, scoring and standard-setting which is needed to conduct and report on a quality assessment. A three- or four-year cycle would allow more time for such work to take place.

Links to International Assessments The most exciting possibility which could be considered is conceptual or statistical links to international programs of assessment. Given the choice of whether to "drive" NAEP downwards into states and local districts or upwards into international assessments, the choice ought to be clear. Given the attention being focused on international economic competitiveness and the role of education in preparing students for further education and for work, it seems to me that our priority ought to be to determine how well our students are doing in comparison to the students in other nations. However, the current International Educational Achievement (IEA) studies are not connected in any way with the NAEP studies, other than that students in local districts are tested twice.

Rather than spend resources and energy trying to drive out state and local testing programs and replace them with NAEP (which NAGB seems bent on doing), it seems to me that it would be much more fruitful to explore formal links with assessment programs internationally. My goal would be to eliminate redundant data collection here, too, by using common data collection techniques or better yet, blending programs together. As it stands now, operating two (or three) national and international assessment programs is not only wasteful of federal and state dollars, it is also wasteful of teacher and student time, since these redundant programs each must be administered locally (in addition to state and local assessment programs).

Utility of NAEP for State and Local Comparisons As a person who strongly supports criterion-referenced interpretations of data (such as "Have or have not students learned what we want them to learn?"), I am not all that interested in knowing whether Michigan scores higher or lower than Indiana or Ohio. However, NAEP can serve both purposes, particularly if the data are reported in terms of percentage of students achieving at satisfactory levels and then comparing these figures across states. I believe that the reporting planned for the 1990 Trial Program will do just this, so that I feel that comparative data will serve a useful purpose.

There will be those who feel satisfied by high comparative performance or angry about low comparative performance on NAEP, but given the data on the standing of the United States in international studies of student performance, I have likened such feelings about a state's relative performance on the NAEP to fighting for the upper or lower deck on the Titanic when the icebergs are in sight. There simply are far more important issues that lie ahead of us as a nation, such as how we assure that every student who leaves high school is work ready in the global competitive market we find ourselves in.

Role of the National Assessment Governing Board I have already touched on this topic in my remarks elsewhere. I have had both positive and negative feelings about the

Board since its inception. The positive ones are due to its broadly representative nature and its independence from the contractor. This, I believe, is a useful feature of the Board which ought to be retained and expanded upon in the future. I have also come to feel a true desire on the part of NAGB for state input on various issues facing National Assessment.

On the negative side, I believe that the Board has not always sought out advice from practitioners before acting, is not sufficiently staffed in order to carry out some of its responsibilities (such as standard setting), is pressured into acting before it should (due in some measure to unrealistically short timelines) and hasn't always been aware of NAEP's past nor given consideration to how it could link (or replace) other federally-funded data collection efforts on student or adult achievement.

However, I have come to realize that a major part of these problems is due to not having persons who are familiar with the state-level educational or technical issues surrounding assessment on the Governing Board itself. I feel that a major change to the Board would be to add two or more assessment directors and curriculum directors from states to the Governing Board. Also, Congress should require that a technical advisory group composed of various constituent groups, including state and local testing directors, be established and that such an advisory group be regularly consulted in the establishment of assessment policy and practice before the Board makes final decisions on either. While standing committees have been used for short periods of time and ad hoc groups have also been used, the on-going advice which a standing group could provide has been missing. Both of these are critical needs which should be taken care of in the re-authorization of NAGB.

Summary

I hope that my remarks are helpful to you as you consider the shape and scope of the National Assessment of Educational Progress of the future. As a supporter of the program, I hope that my remarks will be helpful to you as you discuss these and other important issues. There is a useful and constructive role which a National Assessment can play, but careful thought, particularly by those familiar with the National Assessment of the past as well as the practical and technical issues surrounding assessment, will help to assure that the past is not re-invented nor that useful alternative assessments designs and strategies are not considered. My colleagues and I stand ready to help and we are waiting to be called on. We believe that we have much useful to contribute and ask that we become part of the process of discussion and decision on these and other important issues.

MICHIGAN DEPARTMENT OF EDUCATION

MEMORANDUM

TO: Emerson Elliott
Gordon Ambach
Archie Lapointe

FROM: Edward Roeber

DATE: February 28, 1991

SUBJECT: State Involvement in the Design of National Assessment

Since my return from Washington, D.C. on Monday night, I have given considerable thought to the ETS NAEP Network meeting which was held on February 25, 1991 concerning the National Assessment State Assessment Program. I am very concerned about the nature of the involvement that states are provided concerning the National Assessment project. As you are aware, I feel strongly that states should be directly involved in all phases of the design and conduct of this project. From technical, educational and conceptual points of view, our experiences are invaluable if not unique for agencies seeking to design a national testing program which reaches down to the state or local levels. It was this view which led me to suggest at the April, 1990 Education Information Advisory Committee (EIAC) meeting that participation in the 1992 NAEP State Assessment program be contingent on direct state input into the National Assessment programs.

At the outset, I want to clarify that I am a supporter of National Assessment. As a former staff member, as well as a state testing director with years of experience in involving educators in the design of assessment activities and procedures, however, I am most upset by the manner in which state input is not sought or desired in the design of NAEP's future. I wrote about this on January 22, 1990 and I believe that the situation has not changed fundamentally since then. The February 25 meeting, as well as the activities of of February 26, well illustrates the mere lip service that appears to be given to meaningful state input into the National Assessment design. While the most trivial aspects of NAEP are dwelt on, the longer range issues and design work is carried out without states having a meaningful opportunity to comment.

At the February 25 meeting, activities relating to the 1991 field test of the 1992 program were discussed, as well as further drafts of the 1990 reports were handed out and discussed at length. Virtually the entire meeting was devoted to the passing on of descriptive material from WESTAT, NCES or CCSSO. This material was presented to the large group in a lecture format and little or no opportunity was made available to actually have input from the various testing and curriculum persons in attendance. While it is difficult to have in-depth discussion with 150 people at a meeting, it is not impossible. Given the number of NCES, WESTAT and ETS staff present, it would have been quite possible to break the large group into fifteen to twenty smaller discussion groups and use the staff as recorders. I calculated that the tax payer cost for the meeting was approximately $100,000. For this amount, I question whether any sponsoring agency or attendee received any information that would cause any one to do anything differently. Even the lengthy new drafts of the NAEP reports were available only on the morning of the meeting, so that opportunities for input on these were limited by the sheer volume of material to be reviewed.

In the meantime, a very important list of issues to be focused on at the Tuesday public hearing on the 1994 and 1996 NAEP was passed out. Conveniently, notification about the public hearing was received by us just a few days before the meeting, well after travel plans were set. It was clear to me that once again, lip-service was being given to state involvement into design of NAEP. While you had virtually all NAEP Coordinators and mathematics supervisors in the United States in Washington, D.C. and could have easily devoted an hour to the topic, particularly using small group discussions, no provision was made for us to discuss the issues. Worse still, none of the meeting participants had a chance to frame additional issues or options that could be presented and discussed the next day and there are several which, for whatever reason, were omitted from the list. One such issue is the need for a study of motivation of twelfth graders on National Assessment. If the timing was bad, the January Association of State Assessment Programs/EIAC Assessment Subcommittee meeting could have been used to 1) notify states that a public hearing was to be held and 2) allow input on the issues-options list. As a member of the Congressionally-authorized NAEP Evaluation Panel, I find it interesting that the Evaluation Panel also did not have an opportunity to comment on or add to the list of issues, even though we had a meeting just a few weeks ago.

Ultimately, I left Washington very frustrated and angry about how my time had been abused, as well as completely uncertain as to whether NCES or CCSSO plan to utilize the EIAC Assessment Subcommittee in order to get systematic input on these and other issues related to national tests and examinations. It seems to me that it should be possible to design ways to receive input from the key players in meaningful ways in a timely manner without violating government procedures or causing us to take even more time away from our jobs. Holding meetings at considerable expense to simply present material which could be put in writing is not constructive.

In closing, I request that NCES reconsider the manner in which states are allowed to have input on the National Assessment project. First, the NAEP Network meetings should provide more time to have input on NAEP, at a time and in a manner which allows such input to affect the design of the project. One-way communication (such as has characterized these meetings) can best be handled in memoranda which we can read at home. Utilize the meetings to really hear from the states. Second, when opportunities arise to "discuss" newly-developed items, such as the two proposed state-comparison mock-ups, provide opportunities in the meeting for actual small-group discussion of the items. The presentation and "discussion" of the items in a large-group lecture format should not be disguised as a true opportunity for input on this project.

Finally, when major design and conceptualization issues arise, such as were to be discussed for the 1994 and 1996 NAEP on February 26, why not let the EIAC Assessment

Subcommittee help frame the issues and individually and collectively respond to the issues? To me, **this** is the "involvement" that I meant in April, 1990 and which I had hoped that the new Assessment Subcommittee format would allow. Unfortunately, what I have seen is that states are given minimal opportunities for input and then on only the most trivial aspects of the National Assessment State Assessment project.

Unless this changes, I still feel strongly that states should not participate in the 1992 program. Frankly, I am still not sure whether Michigan should be a participant in 1992. I hope that the states will express their feelings about the design of the National Assessment project and the ways in which states are given opportunities to help shape it. I urge you to reconsider the ways in which states are allowed to have input into this project. While you may be committed to the EIAC Assessment Subcommittee approach for this, are you going to utilize the very mechanism you have established to allow such input? The design of the 1994 and 1996 National Assessments provided the best opportunity so far to test your commitment. How will true involvement of the states in framing the issues and responding to them be carried out?

cc: Association of State Assessment Programs
Association of State Supervisors of Mathematics
David Bayless
Steve Koffler
Gary Phillips
Ramsay Selden
Roberta Stanley
Suzanne Triplett

Chairman **Kildee**. Thank you very much, Doctor, for your testimony. I am going to pick up something you alluded to at the end and throw the question out to all of you for our enlightenment here.

Testing for the most part takes place at the end of an experience or the end of a process. What do we discover in testing that can help us to improve that experience or that process? Testing is generally an omega point. What do we do with alpha, beta, delta, gamma—if I have got my Greek alphabet right—what do we do at the beginning of the process? How can we make testing helpful in the improvement of education?

I think that is probably Congress's greatest concern, to improve education. How can testing, not in isolation be helpful in the improvement of education?

Dr. **Walberg**. I would like to take first crack at your question, sir. I think the implication of your question is extremely constructive, and I think we have to think of testing at various points in the educational process.

For example, Dr. Roeber mentioned the consensus projects that were carried out by the chief State school officers under contract with the National Assessment Governing Board, and this is a way that educators can get together to formulate the test to say, "What are the very best curricula to offer, the best instructional practices?" Then these can be translated into test items.

Then when the test is given, very specific information can be yielded by the National Assessment in terms of where students in a particular State have done well and where they have done poorly.

For example, common fractions versus decimal fractions, and this gives very specific curriculum information to people in State departments of education or we have national organizations such as the National Science Teachers' Association that can look at the strengths and weaknesses of various curricula areas, and that way that there is a constant feedback or interaction between the education process itself and the testing which should jointly enrich one another.

Chairman **Kildee**. Has anyone else got a comment on that?

Dr. **Linn**. Very briefly. I think one of the ways happens all the time in an individual classroom. A good teacher is doing the kind of testing all the time throughout their instruction, and it really becomes merged with the instruction. That is very different than the kind of testing we are talking about when we say a national test or even the National Assessment of Educational Progress.

Where the two sometimes get linked is in what is happening in some States right now where groups of teachers are working together on the type of assessment, and it is really kind of staff development type of work, coming to agree on the kind of performances they are trying to encourage, how they assess those, and how that relates to their instructional program.

Again, however, this is not a mandated national test but rather one that is closer to the classroom.

Chairman **Kildee**. Dr. Roeber?

Dr. **Roeber**. Yes. I think that the testing that would go on—there are different levels of it. There is a classroom level. There is a lot of informal testing that Dr. Linn alluded to. There is also at the State level in our State, at least, assessments that go on that are built on frameworks that are not unlike the National Assessment. But built into that is a lot of assistance down to the classroom teacher to provide them with strategies; for example, to get the teachers together in grades one through four and say, "We have tested these students in fourth grade. Where should they have learned these things, and how do we help them, groups like these students, the next time around and then what do we do with these students that have the needs?"

The National Assessment could do some of that. It is very difficult to have that, but I think we first have to have the attitude that that is appropriate, and when I suggested that, for example, the National Council of Teachers of Mathematics or the Mathematical Science Education Board be given the responsibility of developing strategies like that, it was indicated to me by staff of the Governing Board that that was not an appropriate role for a Federal agency, and that is why I made the comment that it is okay to test people and make them look bad, but not to help them.

It also means that moneys have to be available to groups so that they, in fact, can do that. I don't believe that commissioning the National Council of Teachers of Mathematics to say, "What are the implications of this assessment framework for instruction?" means that you automatically end up with a Federal curriculum, and I think that is the concern.

The area of mathematics is one in which the framework is somewhat forward-looking, and one of the major things that the results this summer when they are reported, I hope, will focus on is that this country really needs to reexamine the whole notion of mathematics instruction.

We have treated it as if it is a special talent that a very few of us have, that we are born with, and the rest of us, it is okay. There is no other area like mathematics. There is no other subject area that you will get grown adults that are presumably well educated that will laugh with one another about being illiterate in this area, and so that one of the fundamental changes that has to occur in mathematics is for us as a nation to say, "It is not acceptable to have adults who are well educated who are mathematically illiterate," and it was this notion that I wanted to try and use the June 6 NAEP release to try and get to.

We are going to have public attention about schools, and one of the things that we need to focus public attention on is the need to change the attitude about mathematics. All students and all ages need it, and it is not a special talent that only a few of us have. It is something that we all can develop, and the nature of mathematics is changing, so that rather than using the NAEP results released on June 6 as an occasion to say, "Schools have failed," we ought to be using it as an occasion to say, "Here is what needs to be done to change and improve mathematics instruction," and I am afraid that that message will not be the one that is featured.

Chairman **KILDEE**. Go ahead.

Dr. **KORETZ**. Mr. Chairman, if I can take a somewhat different tack, I think your question is a very good one because a number of the proposals that are now being floated at the national level would, in fact, take a test that is given at or near the end of high school and use that as a mechanism for driving educational change all the way down through the system.

I think that approach is very risky. It assumes without, I think, a whole lot of evidence that somehow the incentives, the new incentives, if you will pardon the metaphor, will trickle down to the younger grades, and that, therefore, we don't have to worry too much about how that happens.

I am not sure that is true. If you go into a third grade classroom and look at mathematics instruction, for example, the issues that the third grade teacher should be facing are things like, "To what extent should I get rid of traditional drill and replace it with the use of manipulatives? To what degree should I replace individual seatwork with group problem solving?" and any number of questions of that sort.

It is hard to see, in my mind, how a test given in grade twelve will manage to inform those decisions very well unless our goals for education are so narrow that they can be specified in a very lock step sort of way, and in my experience going into classrooms and watching how teachers have responded to State testing given every several years as opposed to once, there is some trickle down there, but it isn't necessarily what you want. It sometimes comes out in the form of pressure for the most routine and boring sorts of drill.

I think the real question, which is often unanswered in the national testing proposals, is just what is going to happen inside schools to change the incentives that teachers and students face? Just simply having a test at the end of the road isn't necessary going to do it, I think.

Chairman KILDEE. I appreciate your answer because I think that is rather fundamental to this issue. The key question is how can we use the tool of assessment or testing, to improve education. That is very important.

I ask that because—and I am going to personalize this a bit. I taught Latin for ten years at the high school level, and I can recall my first couple of years, I just couldn't understand why my students could not master what we call the sequence of tenses, when you use the indicative mood and subjunctive mood and must determine which tense you should use. It was always very clear to me because I had spoken Latin for six years in college. That was the medium of instruction, and I just couldn't understand why my students couldn't master those sequence of tenses.

But through testing I discovered they just really did not understand it very well so I redid the way I presented the material in a very logical way, and I discovered that my slowest student was then able to understand.

So that taught me, then, that by testing my students, I was able to find out they weren't doing well, weren't understanding that well, and therefore I could change my method of presenting and discovered the right method.

If we can link education and assessment in some way with the improvement of education, then that makes us more comfortable in feeling that it is going to achieve some real purpose.

Very often in education we start at the Z point rather than the A point. I think even in schools of choice, which is now on the Hit Parade, that we go to School Z but don't ask what we ought to do with School A which the students are vacating because they don't like what is there. I think we very often have to start at the beginning. We can test, find out what we can do to improve education there, but then apply that in some way.

Do, for example, the National Association of Teachers of Mathematics, when they get together, look at tests and say, "What can we do to improve the method of delivery of mathematics?" Does anyone want to comment on that?

Dr. **WALBERG**. Well, I can say that in general the national professional organizations such as the National Science Teachers' Association, the NCTM—that is National Council of Teachers of Mathematics—are very test conscious because they realize that there is a danger in the test driving everything, and that, I think, is one of the reasons why testing is so salient these days. So many of the professional groups at the national level are looking at these things and translating them through various types of consensus procedures, but it is also being done at the local and the State level, too, and just as you said, and Dr. Linn mentioned, a teacher might give a test and then see if that, as you illustrated, is helping the instruction.

There is a deliberate attempt to link improvement with assessment, and I think it is highly constructive to do that, and that is why I would say that the National Assessment Governing Board would like to make itself as useful as it possibly can. We have many demands; there are many things that we are being asked to do, and I think that many of them are very valid and legitimate, and we have to figure out which kinds of things that we could possibly do best.

For example, we just heard, I think, two points of view. One is that we should give a lot of attention to the—that is, the National Assessment Governing Board attempting to directly improve education, and giving prescriptions on what to be done.

Another group would say, "No, just do the monitoring alone. Let's just have the facts, and leave it up to local educators and people in States to do that."

So I think we have to be very careful about what our purview is, and I think the board has tried to consider that very deliberately.

Mr. **MARTINEZ**. Mr. Chairman, would you yield on that point?

Chairman **KILDEE**. I would be glad to yield.

Mr. **MARTINEZ**. You know, the idea that you are talking about, testing, conjures up in my mind a couple of problems that I have always seen as my kids went through school and as I went through school.

In testing, most tests that I have seen to date actually seem to test how much knowledge you have gained, what you remember about what you have learned in that class. It has nothing to do with how well you will perform any place else. Who is from the RAND Corporation? And you say scores can improve dramatically because you teach to the test. I know that very well because there are real estate classes that they give. All you learn, all you do in those real estate classes is memorize the questions they will probably ask, and memorize the answers. That is what a lot of gaining knowledge in the classroom is, memorizing facts, hoping that that is what the teacher will ask the question on or that is what you will be tested on.

But assessment to me means something very different. It means assessing how well that person will perform in that subject matter regardless of what the test might show, and I don't know how you do that exactly, but there must be a way of doing it.

I mean, I would think that in a case where you ask a question with only one answer, you ask the student in a test to perform a certain task. Write a business letter, you know, rather than ask him, "Is this business letter constructed accurately?"

If you are going to do assessment, I would think that assessment keys to that ability to help because you assess where the weakness is and then you target that weakness by extra instructions or extra help. In the case of the national ability to help, you asked the question, "You can beat up on them because their test scores are not good, but where do you help them?"

I would imagine the Federal Government, which doesn't want to get involved in curricula, could provide moneys that would go to schools that have identified a number of students with weakness or that need extra computer help or materials or whatever it is, or even specialized teacher help, to improve that particular situation.

So my question basically is, how do you distinguish between testing and assessment and how do you achieve what the Chairman has asked for, the ability to assist? And where is the greater ability to assist: after assessment or after testing? I happen to believe it is after assessment because that is where you identify the weaknesses.

Dr. KORETZ. Can I take the first stab at that? There really are a couple of questions there, Congressman Martinez. One of them is how assessment or testing can be used to actually give teachers and administrators information to improve instruction.

I think the answer to that depends on the context, and one of the things that makes me nervous about test based accountability—it made me nervous in the 1980s when States were doing it; it makes me more nervous now when people are suggesting the Federal Government do it—is that it is very hard to know, sitting miles and miles away from a classroom, what information a teacher needs.

There is sort of an implicit model in some accountability models that teachers really know what to do but they are withholding, and if we just set the test at the end and give them incentives, they will finally start producing for us.

I think that is not true. To reduce this to an anecdote, I recently went into a third grade classroom. It was the highest track. It was a school that tracked even in the third grade; highest track math group in a school that scores at the top of the scale, basically, on a State mandated math test, but the class was having clear problems, and I sat down with one student and simply asked him to describe why he was doing certain things that he was doing, and the answer was he had no idea, none.

Since then I have heard parents in that classroom starting to complain that their kids are doing very well but don't know what they are doing.

In a sense what happened is that the State mandated test supplanted—it didn't add to, it supplanted other kinds of information that the teacher should have been using to find out whether kids were learning and what they needed to learn. That kind of displacement, I think, can occur quite often. Deborah Meier, who is the principal at Central Park East High School in New York, said that there once was a day when if parents came in and said, "How does my kid read?" the teacher would say, "Let me show you what kinds of things your kid can read, and let me show you some things that your kid can't read." Now they pull out test scores.

So I think that there is a danger that centrally imposed testing, while it can sometimes have, I think, very desirable effects, can also undermine precisely what we are trying to improve, and that is why as we move into—apparently we are moving into more nationalized testing, I want to see more attention to what mechanism is going to be set up to guarantee that this information actually helps rather than hinders.

Very briefly, one answer to another part of your question is, I think there is now a very widespread desire to increase the content and the range of content of what are included in tests, both at the national and local level, and to include things that are much more along the lines of writing a letter rather than correcting it.

So I think at least on that—we don't always know how to do that in terms of rigorous assessment, but there is clearly a movement afoot to try to do that so I think at least that part of the problem may get better.

Chairman KILDEE. Thank you, Dr. Koretz. Dr. Roeber?

Dr. ROEBER. Yes. I wanted to comment on that because it really touched on several of the—your question, I kind of touched on in my comments, and let me elaborate a little bit.

I mentioned that performance assessment used to be a key feature of the National Assessment program. From 1969 to the mid 1970s, in fact, when I worked at National Assessment we were basically given free rein to design assessments that best assessed the subject area so in career and occupational development, if we wanted to know whether people had life salable skills, we asked them, and if they said they could wait a table, we went to a restaurant and said, "Show us."

In the area of science we gave them complex problems such as Congress is wrestling with such as "How should we generate electricity in the United States?" And then we scored the depth of their thought and the kinds of reasons that they gave.

In citizenship the extent to which they were taking thoughtful stands on complex issues without correct answers. Many of the things that we most value and would, in our own working definitions of what a well educated person is, National Assessment was assessing. They did it within a reasonable budget on an annual testing cycle in the early 1970s. There is not a lengthy institutional memory, however, and so now we are reinventing this and calling it untried and unproven and going slow on it.

The second point is that the real value of the young adult assessment that I referred to was that it forced the designers of the assessment to think about somebody who has gone through all of their formal schooling. What is it that they should have? What should be the residual of that?

And what that really got us to thinking about is, well, we go through school and we learn all those school skills, and you can test those, but what are the real products of that? How should I be able to apply that when I am 29 years old, which I wish I was. You know, how can I apply that; at home, on the job, as a citizen, and all of the roles I would play in society. That is far different than thinking about life as made up of mathematics and science and writing skills. It really talks about the application of those in various life roles.

And finally, I think if you base your assessments on those then the types of assistance that you provide to schools begin to make sense, rather than forcing them into drill and practice for memorization, in fact, what you are setting standards that would cause them to move away from that and talk about is their instruction leading to good learning that stays with people at least as late as age 35.

What NAEP found from that young adult assessments is that a lot of the school stuff that lead to memorization work is clearly forgotten by adults, but those life application skills that 17 year olds were very weak on, adults fortunately are very strong. They probably learned that through the school of hard knocks, unfortunately.

One of the reasons why I would like to see that reinstituted with the National Assessment is I think it could provide some guidance to our high schools or secondary schools as they look at that there is more to life than going on to college, and that even students who go on to college go on and get jobs and have to fulfill a number of life roles, and I think that that would give us a broader perspective on assessment and hopefully a broader perspective on instruction.

Chairman KILDEE. Mr. Goodling.

Mr. GOODLING. Thank you, Mr. Chairman. I guess I would ask you to respond to whether my fears are justified or not. I realize that we are partially responsible in the Federal level that some of these ideas that seem to be driving the wagon at the present time because we are always talking about accountability and quality and excellence. Now the governors and the President are talking about standards and how you measure them and achieve the national goals.

I guess the question I have primarily is, are we doing these things, whether it is choice, whether it is testing, whether it is teacher training, whether it is national curriculum, are we doing them because we don't know what else to do and we are just sort of throwing up our hands and saying, "I don't know what to do." On my side of the aisle, particularly, I think probably nine-tenths of them graduated from a public school, and I think they were successful before they came here. But if you hear them talk, public schools must be the worst institutions in the world.

So I guess my concern is, are we just throwing up our hands when we come up with the idea of maybe a national curriculum, maybe national testing, maybe choice, et cetera, et cetera, rather than clearly thinking through how these different moves may better our education in this country?

I heard Al Shanker, for instance, recently say that he believes there are a lot of good possibilities in relationship to choice if it is not used simply because you say, "Well, I don't know what else to do and therefore apparently choice is the answer to everything."

I keep trying to find out what the administration means, for instance, when they talk about choice. I haven't found that out yet, but perhaps I will be lucky enough one of these days.

Does anybody want to respond to my concern? Is it justified or isn't it?

Dr. **Linn**. I will respond, particularly in the context of using tests in this way rather than choice. I think that too often that is exactly the case. Tests that have been at State levels, at district levels, and now being discussed at the national level too often have been viewed as a cheap fix, as the engine of reform, as opposed to more systemic changes in the educational programs.

One difference that we now have in a proposal for a national examination system as opposed to a national test of the type that is being discussed by Lauren Resnik and Marc Tucker is the notion that you have to start with what the syllabus is and with the educational program and curriculum itself, and that the exam follows from that as opposed to leading it.

Congressman Kildee's earlier comment about his Latin class, I think one of the implicit points that was there was that you had clearly in your mind what the Latin was that you wanted the students to be able to demonstrate, and unless you have that, the tests can lead you astray and in the wrong direction from above.

Dr. **Walberg**. I would like to add to Dr. Linn's points. I think, again, trying to take the National Assessment Governing Board's point of view, I think our goal has been to provide the best information we possibly can and to be somewhat reserved about giving prescriptions.

We heard Dr. Roeber advocating in a sense more prescriptive information, telling what to do, and we heard Dr. Koretz expressing grave reservations about that.

I think the Board's view is that we would like to do the very best we can to provide the information, but what to do with that information might have certain causal uncertainties about it, and each State may be different, each school district, and so what we would like to do is give this information and let local schools, States, Congress, and other agencies draw implications for policy.

We have not been given that specific charge to do that, but I think if we sort of stick to our last, that is one of our best functions, and I think that that is the Board's point of view.

Dr. **Koretz**. Congressman, whether or not your fears are justified I certainly share them. I think a number of the major policy initiatives that are gaining credence today are of the sort that we finally—people are hoping that we finally found some lever that we can pull from afar that is going to make the whole system fall into line.

I don't think there is such a lever. If you will pardon an analogy, it is in some sense analogous to the idea of a vaccination. We are looking for some kind of a vaccination that we can find that will work. If we do, then the government can say, "All right. We will fund development and use of this vaccine, and we are all done."

I think a better analogy for what we are facing is the efforts that have been under way with Federal support for several decades now to reduce heart disease and cancer where the answer turned out to be, unfortunately, there is no vaccine, and there is no antibiotic. What we have to do is somehow arduously do the work to get people to change their lifestyles, to live healthier lives and to avoid some of the risks, things that would put them at risk.

In many ways improving elementary education and secondary education, I think, is analogous to that. We have to change a lot of people's behaviors and expectations all through a very fragmented and large system, and I don't think we can do it with any one lever. I think some of these levers can help, but I think the attraction of them, and the reason why they are in many cases oversold, is because it gives us something we can clearly do from one place, from Washington, from a State capitol. We can just do it, and hope that we can wash our hands of the matter.

I think if we had more realistic expectations for what some of these centralized initiatives could accomplish, we might well accomplish a lot with them, though they wouldn't be the whole answer.

Dr. **Roeber**. Well, I think that we do tend to put a lot of faith in tests and hope, perhaps, as well. One of the things that I have seen from the years of experience that I have had in Michigan, and certainly Pennsylvania has had as many years of testing at the State level as we have there, we tend to view tests as, first of all, a way of determining whether or not there is a problem, perhaps as a motivator for change. We embarrass enough people or we can maybe be more helpful and then just embarrass them, but then also want to rely on the same test, perhaps, to evaluate the effectiveness of the change, and I think we are asking a lot of a test.

We began in the early 1970s what really has become a national phenomenon because we had some researchers in Michigan that were very concerned that testing was being used to indicate that students really could not profit from schools, that all that they could learn was predetermined, almost, before they came there, and the State began a whole train of research on effective schools and also on the process of change.

If you know the characteristics of an effective school, one that is serving well, minority students and others of low economic circumstances, can we go about taking that information into schools that are similar to that but where the performance is lower, and what that research indicated is that the classroom teacher has to be actively engaged in the process of change.

So that assessment in our loosely coupled system—and I say loosely coupled because we have a national test which may or may not affect State policies. We have State assessments which may or may not affect local policies. We have local assessments which may or may not affect school policies, and school policies which may or may not affect the classroom teacher.

In that kind of loosely coupled system to expect that a test from the national, the State, or even the district level, is going to profoundly affect classroom teachers, may be unrealistic, but on the other hand if we look at the change process as starting with teachers, that they have to be actively engaged in a school improvement process that they feel a stake in, then I think that there is a constructive role that tests can play.

Using the health analogy, just finding out that the average American's cholesterol is too high won't necessarily motivate me to change my eating habits, but if somehow somebody can get to me, then we can work on that process of change, and I think that is why I see that it is critical that whenever we test, we try and provide people with information that they can use to improve. That isn't making them do it; that is not holding them responsible that, "You are not going to keep your job unless you do that." It is to provide resources to people and the encouragement because it is very, very difficult to affect classroom teacher behavior, very difficult.

Mr. **Goodling**. I guess one of the concerns that I have, and maybe I was incorrect in doing it, but I always tried to convince teachers that the major purpose for their exam was to determine where they failed to present the material well, whatever it was they wanted the student to learn, and where individual students did poorly and what they might do to help that student become a better student. Then when I hear people talk in terms of national tests, it kind of reminds me of my first experience as a guidance counselor.

I wasn't asked whether I thought grouping in seventh grade would be a good idea. I was told by the superintendent that I would group seventh grade students and so I took the sixth grade records and I eventually decided that this one teacher—they must already be grouped in sixth grade because the one class was just outstanding. Everyone in there did very, very well in all the results that I had so I went to the superintendent and said, "Do you have them grouped already in sixth grade?" He said, "No, why do you ask?" and I told him, and he said, "Well, you have to understand Mrs. So-and-So always teaches to the test."

Well, you know, it sort of took care of my grouping. I thank you very much. It has been very rewarding and very meaningful, the testimony that you have given.

Chairman **Kildee**. Thank you, Mr. Goodling. Mr. Gunderson?

Mr. **Gunderson**. No questions.

Chairman **Kildee**. Let me ask some more questions here, then. Dr. Roeber, we have had 22 years of experience in Michigan in testing. Has the testing process in Michigan been formally utilized to try to improve education, getting back to my first question. Is there a linkage there or has it just been reportorial?

Dr. **Roeber**. No, there has. We have very actively engaged the Michigan Council of Teachers of Math, the Michigan Reading Association. They helped us establish the frameworks on which we base our assessments. They help us generate the assessment materials themselves so that they are as content valid as they can be, and most importantly, once we get done with the assessment, they also help us in interpreting those scores to teachers.

We place a great emphasis, perhaps unique among even State programs, in using the assessments to have a positive impact on instruction. I think if we set the right targets for our students, I am less concerned about teaching to the test or teaching to the assessment. If we are testing the right things in the right ways—I don't mean—I don't want people to practice the actual test items, but I want to make sure that they are teaching the right things, and that is why we have to have that involvement.

It has led, for example, in the area of reading for the reading teachers in Michigan to really consider that that reading assessment is theirs, and although the results are low, that is because we have set some fairly high standards for our students, and as a consequence people feel quite, quite desirous of helping schools improve their instructional programs in reading.

The reading people without State resources have put on literally hundreds or thousands of workshops around the State for classroom teachers suggesting how the model on which the assessment is based can and should affect instruction so that the assessment, when it was implemented after the five years of development, found a very receptive audience because the teachers had already learned about the instructional model that was related to that definition of reading, and we are eagerly looking forward to the first assessment data so that they can see how their students are doing.

Chairman KILDEE. You suggest, Dr. Roeber, that a conceptual or statistical link to international programs of assessment could drive NAEP upwards. Could you go into that further, and how do Michigan students—you mentioned you are not too concerned about how they compare with Indiana but maybe internationally—how do Michigan students compare on assessment with their counterparts in the fourth, eighth, and twelfth grade who go to work in the Nissan plant in Japan?

Dr. ROEBER. I wish I had data for that. In fact, an opportunity was provided to States. Unfortunately, it was too late for us in Michigan to take advantage of it. It occurred after our budget cycle was completed, and although I had pledges of private support, that was made to a previous governor, and for political changes was withdrawn.

Chairman KILDEE. I recall that evening the change took place.

Dr. ROEBER. And as a consequence we couldn't gather the data. What we were proposing to do was to select a Michigan representative sample, augment the IEA data collection, and at the same time administer to that sample a portion of the Michigan reading test so that then we would have been able to predict scores on the IEA scale from the performance of students on the Michigan scale.

The linking of those two scales, then, would have allowed us in future years when the Michigan test was given for us to come up with predicted IEA values, and that would have allowed us to do that. Unfortunately, the funding was not available in time when I had to essentially pull the plug.

Ironically, the current governor is making a lot of noises about wanting to know how Michigan students do in comparison to students in other countries so we may have a chance to fund this program next year. As I understand, over 30 States have indicated an interest in this.

What that would have allowed us to do is to report the proportion of students at a basic or proficient level in this particular Michigan school building in the district, in the State, as well as in the 37 countries that are participating so that I would have gotten the national data equivalent to what I would have gotten out of the

National Assessment Project, but in addition I would have gotten the information from the 36 other countries.

These kinds of things are possible when there is a similar enough definition of the subject matter to be tested and when you have given both tests to at least a sample of students so that then you can make that statistical link. The same sorts of methodology could be used with National Assessment.

There is an alternative that has been done, and that is to take National Assessment items and administer those internationally. I am personally not in favor of that because all we are doing is seeing how other countries do on American standards. The real value of the international program is that we see how American students do on international standards, and I think sometimes our standards are different.

In the area of reading, for example, the U.S. committee on which I serve was predicting that the United States standards in reading would be higher than those in many other developed countries. Our students may not perform as well as those countries, but at least we are expecting more and want more out of our students than these other countries. We will have data, probably by this summer, to see whether or not that prediction was correct.

Chairman KILDEE. Thank you. Dr. Walberg, would you elaborate for the committee on ways that comparisons of States or districts and of schools using NAEP results will help improve education?

Dr. WALBERG. Well, this is a controversial question, but the Board has had a lot of interest on the part of States and local communities that wish to know how they stand on National Assessment standards, and I think because of the release of State comparisons this spring and our performance standards that we will be reporting on in the fall, there will be even more interest in this, and I would say educators are very interested in knowing how they stand on a NAEP scale, but we should not forget that it is not just educators involved. It is State legislators, business people. I think it is quite clear to many Americans these days that even on our own tests, that is, NAEP tests, when they are given in other countries, we have not done well, and so there is a tremendous national concern about our standing.

So if they have to wait for IEA results which may only be given every four years internationally, or they have to wait a long time for NAEP results, they may institute certain changes in their local school or their State or we may institute changes in the Nation, and if we have to wait too long to get the information, then they have to make decisions in the meantime.

So by various ways it would be possible, such as in the old days, to use what were sometimes called released items, that is, items that had already been used and therefore wouldn't hurt to use a second time. That is the fact that it wouldn't hurt the National Assessment itself because those items could be used by local districts.

It would be possible to either reuse them themselves or calibrate it to local tests so that a test in a particular school or district could be calibrated to the NAEP and at any time that they wished, in June or August, they could give a test, calibrate it with the National Assessment, and see where they stand.

One of the great dilemmas in my opinion, sir, of testing is that we don't have a meter stick that is kept in Washington or formerly in Paris where you could compare any student at any time with a universal standard.

NAEP does have that possibility, and I think if we moved in that direction, we would have to be very cautious because we have to protect the trend lines and we have to protect the security of the test, and we also have to maintain professional standards such as not reporting individual students' information.

I think it would have to be done very carefully, but I think the testing, if it was wisely and economically done, efficiently done, has the potential in some of the ways that you have pointed out in giving information on how to improve schools, could be made much more useful today by having these kinds of calibrations so that local school districts and States could make much more use of these things and on a very efficient basis, too, so that we would avoid too much testing so that it enhances rather than interferes with the actual education program.

Chairman **Kildee**. Yes.

Dr. **Koretz**. Could I offer an addendum to that? I think different tests could help in different ways, but I think the National Assessment has one possible way of helping that nothing else does which is it can just simply tell us whether things are getting better or worse.

If you look at the wave of State level education reforms in the 1980s we really don't have as much solid information as we ought to have about whether that extraordinary burst of reform did much, in part because when States use tests to drive instructional systems, they often didn't bother to put in place another, uncorrupted test to see if anything actually got better. They used the test they were using to drive instruction to monitor change, and of course scores went up.

But we do have one sign about what happened in the 1980s, and that is the National Assessment which nobody bothered to teach to, and it didn't budge. That is actually an overstatement. It budged. There was trivial improvement in some areas, slight improvement in others.

So the NAEP managed to tell us, not with as much detail as we would like and not as well as we could get out of a formal program evaluation, but it gives us a pretty good sign that that wave of activity in the 1980s didn't do what we hoped it would do, and it is the only thing we have in the United States that can tell us that.

I think as the next wave of reform takes hold, it would be a shame to lose the ability to monitor whether or not we are getting better which is currently only offered by the NAEP.

Chairman **Kildee**. Yes, Dr. Roeber?

Dr. **Roeber**. Dan referred to the trends during the 1980s. I would reference also the trends during the 1970s. There was substantial improvement in NAEP performance during the 1970s at a time when our nation—perhaps not the States but at least the Nation—paid considerable attention to the achievement of poor and minority students in the United States.

There was both statistical, and most importantly educationally significant gains, on the parts of these students. Unfortunately, while the gap narrowed during the 1970s, that gap did not continue to narrow during the 1980s, and in addition one of the trends is that the performance of students at the highest levels of the NAEP scales, the percentage of students that were capable of performing at that level, has not changed much over that period of time and continues to be quite low so that I think that also illustrates, again—I would agree that it illustrates the value of a monitoring system.

You may not be able to attribute it to a particular policy that took place in a particular State or school district, but you can see whether or not things have changed, and I think that the 1970s was a time when we paid a considerable amount of attention to the achievement of those groups, and we saw significant changes in their performance.

I think that was most heartening to see, and as a reviewer of the 20 year trend in National Assessment that was something that I particularly paid attention to in my comments on that.

Chairman KILDEE. Mr. Goodling.

Mr. GOODLING. Just one last question. A lot of people seem to believe that unless we change teacher training institutions we are not going to change very much. Is there a role in NAEP, for instance, and testing programs that might help us encourage these changes if you think changes are necessary?

And I guess a follow-up to that would be are the teachers trained presently in the use of the results they get from the testing that we now do, or is it just something that we put on the shelf and say it was nice.

Dr. ROEBER. It is very tempting, Mr. Goodling, to suggest that if we simply changed the training of teachers, that somehow that magic bullet would also change student achievement. I think that is clearly a piece of the answer but not the only one.

I mean, most of the teachers that are teaching now will be for quite awhile and so we have to worry also about whether or not those teachers that are currently teaching have the skills that they need.

There are a couple of ways in which I can answer, "No, they don't." First of all, the teachers in most States in the United States are not required to have courses in assessment or measurement to become a teacher. I haven't done a national survey on that, but there have been some regional surveys. I know in Michigan they are not, and I have yet to find a State in which they are.

That is probably good because if they did get a course, it would be the wrong one. What we need to do is not teach teachers how to build tests. I mean, there are testing technicians that can do that, and they clearly need to be trained in that.

What teachers need to know is how to link assessment results to good instruction. That means a formal test, be it a NAEP or a State or commercial test or district test or whatever, but it really means, more than that, the kinds of informal testing that Dr. Linn referred to, and teachers have not been trained in that.

The second thing that they are not being trained in is the ways in which the national curriculum groups are redefining the content areas. Science, the science area, has been significantly redefined in two very different ways, one by the National Science Teachers' Association and the other by the American Association for the Advancement of Sciences.

The typical classroom teacher of science, particularly at the elementary level where you are a teacher of everything, those ways of reform have rolled right on past them. The same thing is happening in mathematics, in reading, in writing and other areas so that while the changes, those changes need to be reflected in teacher training, those are future oriented actions.

What States are wrestling with now is how do we work with the—in the case of Michigan—the 88,000 teachers that are teaching now, most of whom will be teaching five or ten years from now.

We have to try and provide as they are working, struggling, to teach these things to students, we have to try and teach them the assessment methods that they can use, and we also need to teach them how to incorporate these major curricular changes in their instruction; obviously a significant challenge for our State and local school systems.

Chairman KILDEE. Yes, Dr. Linn?

Dr. LINN. I would like to just elaborate on one of the points, and that in the context of the National Council of Teacher of Mathematics curriculum standards.

They provide a different view of what assessment would be like than is too often modelled by tests and so I think that having a better modelling of what that is, when Congressman Martinez was here he referred to the notion that you might learn all those facts and do well on the tests, but not really be able to solve the problem.

Well, that is a good illustration of the type of performance assessment that is trying to be developed now and modelled for teachers that would in mathematics not just have the notion that there is a right answer or how to add two numbers together or use a particular algorithm to always get a particular answer, but really, often it is figuring out what the problem is and approaching it from different points of view and coming up with multiple solutions. That is more important to mathematical thinking than having a factual answer.

Mr. GOODLING. Responding to that last comment, I use the phrase so often that "He has all the answers, but he doesn't understand the question."

Dr. WALBERG. I would like to add to what has been said, Mr. Goodling, by saying that there has been something of a national debate about the content of teacher education programs, and in my view it is very unsettled.

There is one view that our teachers need more fundamental subject matter background. If you are going to teach physics, you should have more physics courses. On the other hand, there are those that believe that you need more courses in pedagogy which would include testing so I think it is rather unsettled, and many universities are still debating this question.

Some of the critics of American education have said that our fundamental technology in education has remained the same since the turn of the century. It is basically teacher presentation, recitation, board work, seat work, homework, and despite all the facts of all the new innovations with computers and media and other techniques that really aren't being used very much.

That may suggest, and I think we have to admit, that testing is a fairly arcane subject. It is not something that—people get Ph.D.'s in it, and it takes a long time. It is very statistical, mathematical, and requires a fair amount of specialization.

So to some extent it could be argued that if we want to look at new innovations in education, that there ought to be testing experts or people who would help the teachers in the school, that would specialize in converting information that they may get from NAEP scales, State scales, locally developed instrumentations, and that this would be hooked up to computers so that both teachers—well, let's say teachers and parents but particularly students themselves would get instantaneous information and feedback as to how well they were doing.

I don't view this as a panacea. I don't think it is proven, but I think it is something that ought to be given a try. Testing could play a very valuable role there because many teacher made tests are amateurishly done. They are not as well done as a technically reliable test.

If we could combine these two to make the test both rigorous in the scientific sense and relevant to the teacher and the student, it may be a way that we could increase American educational productivity.

Chairman KILDEE. Thank you. Dr. Koretz, you make a statement that NAEP is so complex that NAGB needs additional technical expertise to make appropriate decisions. Could you give illustration, or would you call for restructuring of NAGB?

Dr. KORETZ. Well, let me answer the second question first. I think whether NAGB needs to be restructured depends in part on what decisions you make in the reauthorization about what its authority should be.

I think there is a lot to be said for the current structure in that educational goals are not a matter, just of professional expertise. The public ought to have a strong voice, and so having a National Assessment Governing Board that represents diverse public views and diverse constituencies, it seems to me, is very desirable for one end of the decisionmaking process.

But there are, unfortunately, a lot of very difficult technical questions that come up. There are, for example, a lot of technical problems with the current effort to set achievement levels, standards of achievement. There is, to go back to performance assessment, there is a problem in deciding when an assessment is sufficiently different from what we have done already that we need to investigate how it is done before we use it for national reporting.

NAGB does have a resolution of its own suggesting that when changes are really dramatic, there should be two trials before a new assessment is used for national reporting, but even the changes that have been made in the national version of the National Assessment to date, in mathematics, for example, raise some very serious technical problems, and there really isn't a mechanism now for resolving whether the decisions that have been made are the right ones.

The pressure is there to simply produce those assessments in the quickest possible time because people want them now.

For instance, at a recent meeting of the Technical Advisory Board that the Educational Testing Service has put together to advise it on the conduct of the NAEP, several of the members of the advisory committee responded to a presentation about one of the new assessments by saying, "Well, do you, in fact, have any information that this new set of items and new set of tasks actually tells us anything reliable?"

And the answer was really, "No, we don't have much." There is no place in the schedule, and there are no resources set aside for finding out whether that is a good way to do it or whether we should be doing it some other way instead. We simply have to get it out in the field.

Ideally, it seems to me the National Assessment Governing Board, if it is going to have responsibility over technical decisions of that sort, ought to have the ability to decide when they are going far enough in that direction that they need to, say, hold the line against rapid change or call in some kind of technical review, and I think over the next couple of years as we get further into performance assessment we are going to have more and more of those issues coming up.

Chairman **Kildee**. Dr. Linn, you stated that the current trial State assessment is not sufficient to make a determination of the success of full scale assessments for grades four, eight, and twelve. When do you believe there will be sufficient data to come to a sound conclusion?

Dr. **Linn**. Well, that depends upon what Congress decides for 1994, I am afraid.

The point that I tried to make is that by 1992 only a part of the system will have been tested, and in particular trial at twelfth grade will not have been undertaken at that point so I think that if the interest is in eventually having twelfth grade included on a State by State basis that it would be prudent to have a trial at that level before it is instituted operationally. That was the major point.

Chairman **Kildee**. Thank you. You answered my question. That is what I was trying to discover. Thank you very much, Dr. Linn.

Does anyone have any concluding statements? Yes, sir?

Dr. **Walberg**. I would just like to mention one point in response to Dr. Koretz about including other technical experts on the Board.

I think that that merits consideration, but I would like to point out that we do have two full time psychometrists on the staff. NCES has various experts that they have assembled to advise on the National Assessment. The National Academy of Education, a committee co-chaired by Dr. Linn, is giving us evaluation and advice on these matters, and we have felt free to assemble technical panels on highly technical issues to make recommendations to the Board.

Also, it has been very helpful to us to have the Council of the Chief State School Officers has carried out some of the technical work such as the consensus panels, and they have done a lot of this work for us, I think, in a very useful way.

I almost think that right now we have so much technical advice that it is almost too much, and we need a broader perspective that you would get from a more publicly constituted group, although we are certainly seeking all the information and advice that we can.

Chairman **Kildee**. Dr. Roeber?

Dr. **Roeber**. I would like to comment on that because that was one of my remarks, opening remarks.

That is, I believe strongly that the heart of the expertise in the practical matter of carrying out, developing, and using assessments belongs in the States, and the State people have been badly under utilized in the development of NAEP. Yes, there have been a number of ad hoc groups, advisory panels, and so forth, and in fact I think that that is one of the dangers that I try and avoid in running my program. I try and seek out a consistent set of advisors that span both the policy, curriculum, and assessment expertise areas so that all of that is brought together in one place, and in fact people can inform one another of that.

For example, in the policy discussions about the length of the assessment cycle, I am not sure that there is anyone with the expertise to talk about what it takes in the way of time and effort and procedures to adequately develop assessments. The assessments that are being developed at the national level take anywhere from a quarter to a sixth of the time that a State would allocate to the same process, and the procedures nationally are much more complex.

I think by having people with that practical experience on the Board, then when the board is setting policy that can happen.

Second of all, I think there should be a standing advisory committee that can deal with advice to the Board from a technical nature. They have come and they have gone. We were wrestling—I was appointed to, I believe, the second of those two panels. At the start of the second meeting we were dismissed, and the Federal Advisory Committee Act was the reason why we were dismissed.

I don't believe that that procedure should stand in the way of the Board getting ongoing technical and operational advice from the individuals that are most qualified to provide that, and so as you consider the role of the Governing Board for the future, I urge that you carve out a place in which the people that have the practical experience of running programs and developing frameworks can get actively involved in this.

I think it is just a good strategy for the Board, anyway. Thanks.

Chairman **Kildee**. Yes.

Dr. **Koretz**. If I could quickly add to that, I think the notion of having a standing panel of advisors as opposed to ad hoc arrangements would be very desirable.

In fact, in my full testimony that I submitted, I suggested that as one alternative way of getting expertise to the Board. That not only brings expertise to the Board but gets people with technical expertise together in a room to discuss and debate issues that are in front of the National Assessment which I think will improve the quality of the advice they give to the Board.

Chairman **Kildee**. Unless there are further statements, I want to thank all the witnesses who testified. I really believe that we have assembled, really, one of the greatest panels I have had the privilege of listening to. One nice thing about auditing a course is that you don't have to take a final test.

[Laughter.]

Chairman KILDEE. Your testimony has added a great deal to this subcommittee's understanding of NAEP and the educational testing issue. You raised a number of important questions. I would like to submit, if you would, please, additional questions to each one of you for the record.

In addition to that we will keep the record open for two weeks for other submissions, and again I thank you, and with that the subcommittee will stand adjourned.

[Whereupon, at 10:55 a.m., the subcommittee was adjourned subject to the call of the Chair.]

[Additional material submitted for the record follows.]

MAJORITY MEMBERS
DALE E. KILDEE, MICHIGAN, CHAIRMAN
GEORGE MILLER, CALIFORNIA
PAT WILLIAMS, MONTANA
MATTHEW G. MARTINEZ, CALIFORNIA
CARL C. PERKINS, KENTUCKY
CHARLES A. HAYES, ILLINOIS
THOMAS C. SAWYER, OHIO
MAJOR R. OWENS, NEW YORK
NITA M. LOWEY, NEW YORK
JOLENE UNSOELD, WASHINGTON
WILLIAM J. JEFFERSON, LOUISIANA
JOHN F. REED, RHODE ISLAND
TIM ROEMER, INDIANA
CRAIG A. WASHINGTON, TEXAS
PATSY T. MINK, HAWAII
JAIME B. FUSTER, PUERTO RICO
WILLIAM D. FORD, MICHIGAN (EX OFFICIO)

(202) 225-4388

SUSAN A. WILHELM, STAFF DIRECTOR

COMMITTEE ON EDUCATION AND LABOR
U.S. HOUSE OF REPRESENTATIVES
320 CANNON HOUSE OFFICE BUILDING
WASHINGTON, DC 20515

SUBCOMMITTEE ON ELEMENTARY, SECONDARY,
AND VOCATIONAL EDUCATION

MINORITY MEMBERS
WILLIAM F. GOODLING, PENNSYLVANIA
SCOTT L. KLUG, WISCONSIN
THOMAS E. PETRI, WISCONSIN
MARGE ROUKEMA, NEW JERSEY
STEVE GUNDERSON, WISCONSIN
STEVE BARTLETT, TEXAS
PAUL B. HENRY, MICHIGAN
SUSAN MOLINARI, NEW YORK
JOHN A. BOEHNER, OHIO

March 15, 1991

Edward D. Roeber, Supervisor
Michigan Educational Assessment
 Program
P.O. Box 30008
Lansing, MI 48909

Dear Mr. Roeber:

Thank you very much for the excellent testimony you presented at the March 13, 1991 hearing of the Subcommittee on Elementary, Secondary, and Vocational Education. The Subcommittee received a wealth of valuable information which will greatly assist our efforts to review the National Assessment of Education Progress.

Enclosed please find the additional questions which I indicated I would be asking you to respond to for the record. Due to the great interest in the issue, I would appreciate your assistance in returning the answered questions by March 22, 1991 so the hearing record can be printed quickly.

Again, thank you for your valuable testimony and your assistance in expediting the return of your answers to the additional questions.

Sincerely,

Dale E. Kildee
Chairman

dt
enclosure

QUESTIONS FOR DR. ROEBER

QUESTIONS

1) YOU STATE IN YOUR TESTIMONY THAT A CONCEPTUAL OR STATISTICAL LINK TO INTERNATIONAL PROGRAMS OF ASSESSMENT COULD DRIVE NAEP UPWARDS. WOULD YOU PLEASE ELABORATE ON THIS STATEMENT?

2) SOME EDUCATIONAL EXPERTS BELIEVE THAT NAEP CANNOT BE BOTH AN ACCOUNTABILITY TOOL AND A INSTRUCTIONAL TOOL. THEY SUGGEST THAT NAEP CAN EITHER BE USED TO TELL POLICYMAKERS WHAT KIDS KNOW OR IT CAN BE USED AS AN INSTRUCTIONAL AID TO HELP TEACHERS AND KIDS IMPROVE. CAN NAEP OR ANY TEST DO BOTH?

3) DO WE RUN THE RISK OF OVERBURDENING OUR STUDENTS AND TEACHERS WITH TESTING? HOW MUCH TIME IS ALREADY CONSUMED BY TESTING IN THE AVERAGE STUDENT'S ACADEMIC YEAR? HOW MUCH MORE TIME MIGHT BE TAKEN UP WITH TESTING IF WE MOVED TO A SYSTEM OF NATIONAL ASSESSMENT OF ALL STUDENTS?

4) TO WHAT EXTENT DO WE ALREADY HAVE "NATIONAL TESTS," THROUGH THE MAJOR COMMERCIALLY PREPARED ACHIEVEMENT TESTS THAT ARE ADMINISTERED IN SCHOOLS THROUGHOUT THE NATION; THE COLLEGE ENTRANCE EXAMINATION SUCH AS THE SCHOLASTIC ACHIEVEMENT TEST AND THE AMERICAN COLLEGE TESTING PROGRAM, OR THE FEDERALLY FUNDED NATIONAL ASSESSMENT OF EDUCATION PROGRESS? IS THERE SUFFICIENT CONSENSUS ON WHAT CHILDREN SHOULD KNOW TO CREATE A VALID TEST OR SYSTEM OF ASSESSMENT?

5) IN YOUR OPINION, IS THE USE OF NAEP AS A "THERMOMETER" TO MEASURE THE STATUS OF U.S. EDUCATION COMPATIBLE WITH AN EXPANDED USE OF NAEP AS A HIGH STAKES NATIONAL EXAMINATION COMPARING PERFORMANCE OF INDIVIDUAL SCHOOLS OR INDIVIDUAL STUDENTS?

STATE OF MICHIGAN
DEPARTMENT OF EDUCATION
P.O. Box 30008
Lansing, Michigan 48909

DONALD L. BEMIS
Superintendent
of Public Instruction

STATE BOARD OF EDUCATION
DOROTHY BEARDMORE
President
GUMECINDO SALAS
Vice President
BARBARA DUMOUCHELLE
Secretary
MARILYN F. LUNDY
Treasurer
CHERRY H. JACOBUS
NASBE Delegate
DICK DeVOS
BARBARA ROBERTS MASON
ANNETTA MILLER

GOV. JOHN M. ENGLER
Ex Officio

March 21, 1991

The Honorable Dale E. Kildee
Chairman, Subcommittee on Elementary,
Secondary and Vocational Education
Committee on Education and Labor
U.S. House of Representatives
320 Cannon House Office Building
Washington, D.C. 20515

Dear Congressman Kildee:

Thank you for your kind remarks regarding my testimony to the Subcommittee on the National Assessment of Educational Progress (NAEP). I have responded to your additional questions on the attachment.

I hope that these comments will be helpful to you and the other Subcommittee members. If they or staff have any questions about this, please feel free to contact me again at 517/373-8393. Thank you again for the opportunity of sharing the state perspective on NAEP.

Sincerely,

Ed Roeber

Edward D. Roeber
Supervisor
Michigan Educational
Assessment Program

cc: Ms. Roberta Stanley
Dr. Teressa Staten

Responses to Questions

Edward Roeber

March 20, 1991

1. Explain the conceptual or statistical links between assessment programs.

The purpose in raising the possibility of linking examinations was to provide an alternative to the need for additional, overlapping or redundant tests or examinations. If Test A and Test B both cover the same content area (e.g., reading) in a similar manner, it will be possible to predict scores on one test from the other. This statistical linking is accomplished by having the same set of students take both Test A and Test B. Then the scores from the two tests can be placed on the same scale. In the future, only Test A **or** Test B **but not both** need be given; the scores on the one test not given can be predicted from the one that was given.

Conceptual linking is carried out when the content of two tests are so similar that the two exams are essentially measuring the same thing. This rarely happens, so that statistical linking is almost always needed.

2. Can NAEP be used both as an accountability tool and an instructional tool?

This dilemma is faced all the time at the state and local level. Rarely are tests given which we do not try to get educators to directly use to improve teaching and learning. At the national level, National Assessment is the best such example that we have of an assessment not currently designed to directly impact instruction or to hold state or local educators accountable for performance.

If NAEP scores are used to judge state or local school systems, NAEP as an accountability tool will be enhanced. However, the more a test is used to promote accountability, the less its value for instructional improvement. Educators may become so concerned about student performance that instead of using the assessment as an indicator of the need to review and improve the instructional program, they may simply teach the areas of student weakness, without providing the appropriate, integrated and long-term instruction which students really need in order to retain the concepts.

It is not impossible for a test to serve both purposes, but the more of one purpose it does, the less of the other that may be feasible.

3. Are schools overburdened with too much testing and will new tests add to this burden?

I do not think you could find many educators who believe that there is too little testing! One major reason for this, I suspect, is that much external testing is not tied strongly to instruction. Hence, any such data will be viewed as extraneous and not needed. Another assessment will simply add more testing time to the burden and is unlikely to change educators' opinions about testing.

The amount of time spent on testing varies from district to district and state to state, as well as from student to student. Those students who are involved in special programs (i.e., special education or compensatory education) receive extra tests on top of those already routinely given to all students. It would not be surprising to find that some students will spend upwards of twenty hours of time yearly on tests.

New national examinations, depending on their nature, could add as little as an hour or two to the total, or in some of the more elaborate schemes, might require weeks or months of preparation time as well as considerable time for each student to carry out their performance.

4. Do we already have "national" tests? Is there sufficient consensus on what children should know in order to build a national test?

The answer to this is both "yes" and "no." Yes, there are various tests which are given nationally, but no, these are not given to all students at a common grade level or are not given to respresentative sample of students. In some cases, such as the ACT and SAT, they are taken only by a self-selected sample of students, so that the data is not representative of the state or the district. In the case of the commercially-available tests, serious questions have been raised regarding the representativeness of the norms groups, since this is still a voluntary option on the part of the districts and schools selected in the sample.

NAEP is one of the nationally-representative assessments which are given in the U.S. Others include the achievement section of the National Educational Longitudinal Study (NELS) and the International Educational Achievement (IEA) studies. However, the NAEP assessment is the most comprehensive of these assessment programs.

I do believe that there is (or can be) sufficient consensus on what our students need to know in order to create a valid system of assessment. In fact, I would worry more about various states and local districts setting different systems in place. However, the feasibility and desirability of arriving at a consensus of what we want our students to know does not mean that national tests are needed or desired.

5. Is using NAEP as a "thermometer" compatible with expanding NAEP into a high stakes national examination for schools or students?

This is similar to question two. Do we expect a thermometer to heat the water if it finds the temperature too cold? Does it still reflect accurately what the water temperature was (or is)? Would another thermometer find the water temperature unchanged? One danger in using the same test to measure water temperature and to "heat the water," as mentioned in the response to question two, is that the changes in student performance may be more attributable to pressure put on schools by the particular test and not represent true student learning which would generalize to other comparable measures. The comparable tests for which schools are not teaching, may not show the same improvement. Hence, the value of the thermometer may be destroyed by using it to "heat the water."

MAJORITY MEMBERS
DALE E KILDEE MICHIGAN CHAIRMAN

GEORGE MILLER CALIFORNIA
PAT WILLIAMS MONTANA
MATTHEW G MARTINEZ CALIFORNIA
CARL C PERKINS KENTUCKY
CHARLES A HAYES ILLINOIS
THOMAS C SAWYER OHIO
MAJOR R OWENS NEW YORK
NITA M LOWEY NEW YORK
JOLENE UNSOELD WASHINGTON
WILLIAM J JEFFERSON LOUISIANA
JOHN F REED RHODE ISLAND
TIM ROEMER INDIANA
CRAIG A WASHINGTON TEXAS
PATSY T MINK HAWAII
JAIME B FUSTER PUERTO RICO
WILLIAM D FORD MICHIGAN (EX OFFICIO)

(202) 225-4388

SUSAN A WILHELM STAFF DIRECTOR

COMMITTEE ON EDUCATION AND LABOR
U.S. HOUSE OF REPRESENTATIVES
320 CANNON HOUSE OFFICE BUILDING
WASHINGTON DC 20515

SUBCOMMITTEE ON ELEMENTARY, SECONDARY
AND VOCATIONAL EDUCATION

MINORITY MEMBERS
WILLIAM F GOODLING PENNSYLVANIA
SCOTT L KLUG WISCONSIN
THOMAS E PETRI WISCONSIN
MARGE ROUKEMA NEW JERSEY
STEVE GUNDERSON WISCONSIN
STEVE BARTLETT TEXAS
PAUL B HENRY MICHIGAN
SUSAN MOLINARI NEW YORK
JOHN A BOEHNER OHIO

March 15, 1991

Herbert J. Walberg, Board Member
National Assessment Governing Board
1100 L Street, NW. Suite 7322
Wasington, D.C. 20005

Dear Dr. Walberg:

Thank you very much for the excellent testimony you presented at the March 13, 1991 hearing of the Subcommittee on Elementary, Secondary, and Vocational Education. The Subcommittee received a wealth of valuable information which will greatly assist our efforts to review the National Assessment of Education Progress.

Enclosed please find the additional questions which I indicated I would be asking you to respond to for the record. Due to the great interest in the issue, I would appreciate your assistance in returning the answered questions by March 22, 1991 so the hearing record can be printed quickly.

Again, thank you for your valuable testimony and your assistance in expediting the return of your answers to the additional questions.

Sincerely,

Dale E. Kildee
Chairman

dt
enclosure

QUESTIONS FOR DR. WALBERG

QUESTIONS

1) IN YOUR OPINION, IS THE USE OF NAEP AS A "THERMOMETER" TO MEASURE THE STATUS OF U.S. EDUCATION COMPATIBLE WITH AN EXPANDED USE OF NAEP AS A HIGH STAKES NATIONAL EXAMINATION COMPARING PERFORMANCE OF INDIVIDUAL SCHOOLS OR INDIVIDUAL STUDENTS?

2) DOES THE NATIONAL ASSESSMENT GOVERNING BOARD SEEK MORE THAN THE AUTHORITY TO DO PLANNING FOR STATE-LEVEL ASSESSMENTS BEYOND 1992? GIVEN THE ABSENCE OF ANY EVALUATION OF THE STATE-LEVEL ASSESSMENT EFFORT TO DATE, IS THERE ANY REASON FOR THE CONGRESS, AT THIS TIME, TO CONSIDER PERMANENTLY AUTHORIZING STATE ASSESSMENTS?

3) PLEASE DESCRIBE THE PROCESS THROUGH WHICH THE NATIONAL ASSESSMENT GOVERNING BOARD SET ITS BASIC, PROFICIENT, AND ADVANCE THRESHOLDS FOR PERFORMANCE ON THE MATH NAEP ASSESSMENT. HOW DID THE GOVERNING BOARD MEASURE THE RELATIVE DIFFICULTY OF INDIVIDUAL QUESTIONS ON THE NAEP ASSESSMENT? HOW SHOULD THESE THRESHOLDS BE INTERPRETED AND USED BY POLICYMAKERS?

4) PLEASE DESCRIBE THE CONSENSUS PROCESS WHICH NAGB GOES THROUGH TO DETERMINE TEST SPECIFICATIONS.

5) WHAT RESEARCH EVIDENCE IS THERE THAT FORMAL STUDENT ASSESSMENTS AND THEIR RESULTS ACTUALLY LEAD TO CHANGES IN THE BEHAVIOR OF STUDENTS AND TEACHERS? ARE THOSE CHANGES PRODUCTIVE OR ARE THEY SOMETIMES COUNTERPRODUCTIVE?

Dr. Walberg
Questions
Page Two

6) HAS NAGB OR NAEP ESTIMATED WHAT IT WOULD COST FOR A NATIONAL TEST GIVEN TO EACH CHILD?

7) WITH THE INCREASED ATTENTION AT THE STATE AND LOCAL LEVEL IN SO-CALLED AUTHENTIC ASSESSMENT, THAT IS, TESTS OF WHAT STUDENTS ACTUALLY KNOW AND CAN DO, TO WHAT EXTENT HAS NAEP INCORPORATED ANY OF THESE NEW ASSESSMENT TECHNIQUES? HOW DO MINORITY AND DISADVANTAGE CHILDREN PERFORM ON AUTHENTIC ASSESSMENT?

8) PLEASE DESCRIBE THE PROCESS NAGB HAS ESTABLISHED WHICH ENSURES THAT QUESTIONS ARE BIAS FREE?

9) HOW DOES LACK OF MOTIVATION AFFECT THE ACHIEVEMENT LEVEL OF STUDENTS, PARTICULARLY HIGH SCHOOL SENIORS? PLEASE DESCRIBE WHAT NAGB IS DOING TO CONFRONT THE VERY REAL ISSUE OF THE LACK OF STUDENT MOTIVATION?

THE NATION'S REPORT CARD

National Assessment Governing Board

National Assessment of Educational Progress

March 22, 1991

The Honorable Dale E. Kildee
Chairman
Subcommittee on Elementary, Secondary and
Vocational Education
Committee on Education and Labor
U.S. House of Representatives
Washington, D.C. 20515

Dear Mr. Chairman:

Thank you very much for the opportunity to testify before your subcommittee on March 13. Enclosed are my responses to the additional questions you forwarded to me.

Please feel free to call upon me, other members of the Board, or NAGB staff if there is any way we can be of assistance.

Sincerely,
Herbert H Walberg
Herbert J. Walberg
Board Member

Enclosures

1100 L Street, N.W
Suite 7322
Washington, D.C. 20005-4013
(202) 357-6938

RESPONSES TO
ADDITIONAL QUESTIONS FOR DR. WALBERG
FROM CHAIRMAN KILDEE
SUBCOMMITTEE ON ELEMENTARY, SECONDARY AND VOCATIONAL EDUCATION

1. In your opinion, is the use of NAEP as a "thermometer" to measure the status of U.S. education compatible with an expanded use of NAEP as a high stakes national examination comparing performance of individual schools or individual students?

Answer:

The National Assessment Governing Board believes that NAEP is our best measure of the status of U.S. education. The Board believes that the action by the Congress to authorize trial state assessments in 1990 and 1992 prudently expands NAEP's ability to monitor the progress of U.S. education without "raising the stakes."

This expansion will help monitor state performance. It permits comparison among states, among regions and to the nation of overall student results and student performance on various subscales; establishes base-line data for states to use in measuring progress over time; and provides useful background information on variables related to educational effectiveness (e.g. teacher training and parent involvement). But this expansion has none of the characteristics of a state level program evaluation or accountability mechanism, as some have asserted. However, the assessment data can help states identify potential strengths and weaknesses in their programs that could then be used to guide their own program evaluation efforts.

The Board has adopted a policy, described in detail in my testimony, recommending amendments to the NAEP legislation that would permit the use of NAEP and NAEP items by states and districts for reporting below the state level. Such uses of NAEP would be at local option and cost, subject to uniform requirements for test administration, security and reporting, and would be separate from the federal NAEP program. The policy further stipulates that public reporting of individual student scores would be prohibited.

In my opinion, such local uses of NAEP would not be incompatible with its use as a monitor of national and state education progress.

2. Does the National Assessment Governing Board seek more than the authority to do planning for state-level assessments beyond 1992? Given the absence of any evaluation of the state-level assessment effort to date, is there any reason for the Congress, at this time, to consider permanently authorizing state assessments?

Answer:

The Board's position is that, unless the congressionally required independent evaluation demonstrates that they cannot yield valid, reliable state representative data, state-level assessments should be made a regular part of NAEP.

However, with respect to the 1994 and 1996 assessments, there exists an issue of timing. The award for NAEP operations for 1994 and 1996 assessments is scheduled for FY 1992. The current NAEP statute, which expires in FY 1993, authorizes trial state assessments only in 1990 and 1992. Thus, for the 1994 and 1996 assessments, there is a mismatch between the award cycle and the NAEP authorization cycle.

Providing explicit authority now for planning for state-level assessments in 1994 and 1996 would allow for needed preparations which, if indicated by negative evaluations of the trial state assessment, could be cancelled. However, if evaluations indicate that state-level assessments are successful but are not authorized before NAEP reauthorization in FY 1994, there may be insufficient time to continue them until the 1996 assessment at the earliest.

Knowing that this question is a concern of the Congress, the National Academy of Education (NAE) plans to issue in March 1991 an interim report on the 1990 state trial assessment and a final report in September 1991. These reports, as well as the June 6, 1991 release of the 1990 mathematics assessment results, should be useful in informing the Congress' decision.

Dr. Robert L. Linn, co-chair of the NAE panel that is conducting the independent evaluation of the NAEP trial state assessment, provided a preview of the NAE evaluation in testimony before your subcommittee on March 13, 1991. In discussing the evaluation, he said that "...I can say that the panel has not identified major flaws in the sampling or administration of the Trial State Assessment that should preclude the release of results as scheduled in June. I can also report that the initial experience was sufficiently positive and without adverse effects on the national assessment that the panel will recommend continuation of an expanded trial in 1994."

It appears that NAEP will be an important component for reporting on progress toward national educational goals. In addition, 40 states have already indicated their intention to participate in the second trial state assessment in 1992, and we estimate that the total will be 43-45 states. This indicates a high degree of interest on the part of the states; we believe that this interest reflects confidence in the first trial state assessment and the plans for the second, and is likely to continue.

Thus, timing of the new NAEP award, the likely positive recommendation of the NAE panel, the continued high level of voluntary participation by the states, and the need for state representative data for reporting progress on national education goals, all seem to argue in favor of the Congress considering action to permit NAEP to collect and report state-representative data for the 1994 and 1996 assessments.

3. Please describe the process through which the National Assessment Governing Board set its Basic, Proficient and Advanced thresholds for performance on the math NAEP assessment. How did the Board measure the relative difficulty of individual questions on the NAEP assessment? How should these thresholds be interpreted and used by policymakers?

Answer:

The Board is in the process of writing a summary report on its process of setting achievement levels of basic, proficient, and advanced in mathematics on the 1990 national assessment. That report is still in the draft stage and a copy will be sent to you as soon as it is completed. The following is a brief synopsis of the process followed in setting the achievement levels.

At its December 1989 meeting, the Board adopted a motion to proceed in principle with the setting of achievement levels for the 1990 mathematics assessment. A public hearing was held in Washington in January of 1990 at which nearly 25 individuals and organizations presented oral or written testimony. Subsequent to this hearing, the Board adopted a plan to set the levels and proceeded to implement that plan over the course of the next several months. Please see Attachment 1, a copy of a memorandum from Roy Truby, NAGB Executive Director, addressed to Governor Roy Romer, dated December 19, 1990, which summarizes the process from May 11, 1990 until December of 1990.

Subsequent to the activities detailed in the memorandum to Governor Romer, the Board decided to collect additional data on the draft mathematics achievement levels. This decision was, in large part, a result of comments received during the Board's third public hearing on achievement levels held in January of 1991. The additional data will come from a replication/validation activity that the Board now has underway. The replication/validation activity will occur at four regional sites (Michigan, Florida, Connecticut, and California). We were advised that more classroom teachers should be involved and have acted accordingly. At each site approximately 60 classroom mathematics teachers and approximately 12 non-educators will participate. The replication/validation activity will result in ratings from almost 300 individuals, about 240 of whom are classroom teachers.

Following the conclusion of these four regional meetings in April, the draft descriptions for each level may be modified. The Board intends to take final action on setting the achievement levels at its meeting in Washington on May 10-11, 1991.

The Board used two means to judge the relative difficulty of various individual questions on the NAEP assessment: (1) by asking panels of expert judges to rate independently the difficulty of each question at each of the three grades tested by NAEP, and (2) by having the expert panels examine data showing how many students at each grade successfully answered the questions on the assessment.

Achievement levels in mathematics will provide policymakers for the first time with an informed judgment about how much of the content of the mathematics assessment students at various levels of achievement and at various grade levels should know. Prior to this point, policymakers, as well as others, including parents, and the general public were only given information about what students did or did not know, without any benchmarks or reference points for judging the adequacy of this performance.

In addition, having three levels (basic, proficient and advanced) will permit policymakers to monitor the distribution of performance, not just average performance. This will assure attention to both high and low performing student populations and permit setting goals and monitoring progress within each achievement level at each grade.

4. Please describe the consensus process which NAGB goes through to determine test specifications.

Answer:

The consensus process that the Board uses to determine the content and test specifications for each assessment is comprehensive, broad, and inclusive. It begins with the Board issuing an RFP detailing the steps, processes, and procedures potential bidders must execute to assist the Board in accomplishing its statutory task. Should the subcommittee want, I would be pleased to send you copies of the RFP work statement and/or any of the documents produced as a result of the consensus process.

Again, and very briefly, the successful contractor convenes both a steering committee and a planning committee of experts and knowledgeable lay citizens to supervise and guide the consensus process. Routinely, public hearings are held, expert testimony is collected, interested parties are consulted, and reviews of recommendations are conducted prior to the Board's final action to approve the content framework and the test specifications.

The result is a very detailed and extensive description of what should be included in each assessment. This description is used to develop the actual items that will be used in each test. As an example, I am attaching a copy of the Mathematics Objectives for the 1990 Assessment (Attachment 2).

5. What research evidence is there that formal student assessments and their results actually lead to changes in the behavior of students and teachers? Are those changes productive or are they sometimes counterproductive?

Answer:

There is a great deal of research evidence, as well as anecdotal evidence that, when properly used, formal assessments lead to changes in behavior of students and teachers. We know, for example, when teachers are trained to use assessments as diagnostic tools, that classroom instruction can become more individualized and that individualized instructional strategies aimed at reinforcing student strengths and overcoming weaknesses frequently result in improved student performance.

However, under the present requirements in P.L. 100-297, which prohibit reporting national assessment data on the individual student, school, or district level, it is highly unlikely that any particular teacher or student will modify behavior based on NAEP results.

6. Has NAGB or NAEP estimated what it would cost for a national test given to each child?

Answer:

NAGB has never maintained that NAEP should be administered to each child. Therefore, NAGB has not estimated the cost of converting NAEP from an assessment based on a representative sample of students in grades four, eight and twelve to a test of all students in those grades.

7. With the increased attention at the state and local level in so-called authentic assessment, that is tests of what students actually know and can do, to what extent has NAEP incorporated any of these new assessment techniques? How do minority and disadvantaged children perform on authentic assessments?

Answer:

The term "authentic assessments" generally refers to testing methodologies in which the examinee produces responses, as contrasted with choosing a response in a multiple choice test item. Both types of items, if properly designed and scored, can be useful in describing what students know and can do. Assessments that use both types of items increase the range of skills and knowledge that can be assessed.

NAEP has much experience with authentic assessment. As early as the 1970's, NAEP developed innovative approaches to assessing music and the visual arts, which included observations and evaluations of individual and group student performance. NAEP writing assessments involve solely the production of student writing samples. Approximately 40% of the 1992 reading assessment is devoted to student produced responses. The 1992 reading assessment also includes a study that involves direct student interviews and portfolios. The 1990 mathematics assessment included items to assess higher order thinking skills, computer skills, use of measurement instruments, and calculator applications. Planning for the 1994 science assessment is now underway, and it is likely that it will include a balance of multiple choice and student performance items.

We do not know conclusively how minority and disadvantaged students will perform on "authentic" type NAEP assessment items. However, NAEP trend data for reading, mathematics and science from the mid-1970's through the mid-to-late 1980's indicate a significant reduction in the gap between the performance of white and minority students which holds true across ages 9, 13 and 17. The minority gains at the higher ages, involving assessments of increasingly complex subject matter, could be a positive indicator for performance on "authentic" items, but this is not certain. NAGB will be following this issue closely. Clearly, state-representative data on this issue will provide state officials with information necessary to determine whether state policies and programs need evaluation or revision.

8. Please describe the process NAGB has established which ensures that questions are bias free?

Answer:

The Board has developed a comprehensive policy that is used in the development and review of cognitive items for each assessment. A copy of the policy is attached (Attachment 3).

9. How does lack of motivation affect the achievement level of students, particularly high school seniors? Please describe what NAGB is doing to confront the very real issue of the lack of student motivation?

Answer:

The Board is concerned that high school seniors selected to participate in NAEP may not have incentives to put forth their greatest effort. For example, these students know that their individual NAEP results will not be reported to their schools or their families and will have no effect on their grades or graduation prospects. Currently, there is no evidence that a problem exists, but the Board has asked the NAEP Technical Review Panel (TRP), established by the National Center for Education Statistics, to conduct research on this issue. The TRP is developing recommendations for a study design that would be conducted as a part of the 1992 assessments.

Attachment 1

THE NATION'S REPORT CARD naep

National Assessment Governing Board

National Assessment of Educational Progress

MEMORANDUM

TO : Gov. Roy Romer, Chairman
National Education Goals Panel

FROM : Roy Truby, Executive Director Roy Truby
National Assessment Governing Board

DATE : December 19, 1990

SUBJECT : Summary of Achievement Level Setting Process

On May 11, 1990 the National Assessment Governing Board (NAGB) unanimously approved a plan to set the first standards for student achievement on the National Assessment of Educational Progress (NAEP), which tests a representative cross-section of students in grades 4, 8, and 12. The plan had been under discussion and review since the summer of 1989; it was a modified version of a proposal presented to the Board in December 1989 which was the subject of a lengthy public hearing last January.

Under the plan NAGB will define three levels of achievement --basic, proficient, and advanced--for each grade and subject tested by the National Assessment. It then will report what proportion of students have reached each level both nationwide and in participating states. The system is in contrast to NAEP's past practice of simply describing how students perform with no reference to standards of how well they ought to do.

The Board was authorized to identify "appropriate achievement goals" on NAEP as part of the legislation that created NAGB in 1988, long before any national education goals were formulated. In adopting its policy, however, the Board kept clearly in mind the need to help track progress toward Goal Three, set by President Bush and the nation's governors, that "by the year 2000 American students will leave grades 4, 8, and 12 having demonstrated competency in challenging subject matter, including English, mathematics, science, history, and geography." In fact, the phrase, "demonstrating competency over challenging subject matter," has been incorporated as the main defining language of the Board's description of the proficient level for each grade.

To us the division of responsibilities between NAGB and the National Education Goals Panel seems clear. NAGB, through its processes of wide consultation with experts and the public, can define in considerable detail the content of what different levels of achievement ought to be. The Goals Panel can then set numerical goals and targets for the percentage of students that should be at

1100 L Street, N.W.
Suite 7322
Washington, D.C. 20005-4013

each defined level by the year 2000. Alternatively, the panel might set targets and a timetable for the increase in the proportion of students at or above each achievement level. Thus, together, NAGB and the National Education Goals Panel can set goals and targets to give substantive meaning to Goal Three.

NAGB is setting achievement levels first on the 1990 National Assessment of mathematics, which is providing the first state-level data on school performance that is based on comparable, representative samples of students. The Board expects to define achievement levels on the new NAEP tests of reading and writing for 1992 and in science, U.S. history, and geography for 1994.

Process Used to Set Mathematics Standards

NAGB understands well that the judgmental process used to define what achievement ought to be should include a wide range of interests and perspectives. Since this standard-setting effort is perhaps the largest and most important ever in American education, the Board felt it must be open to public scrutiny and input.

The following is a summary of the process for setting math achievement levels:

Appointment of Advisory Panel--June 1990

NAGB appointed a panel of 63 judges. About 70 percent of the panel members were educators, representing subject area teachers, college math instructors, principals, and state and district curriculum specialists. About 30 percent were non-educators--employers, civic group representatives, and interested citizens. About 20 percent of the panel were minority group members. Attention was given also to gender and geographical representation. Panelists came from schools from New York to California, from the inner-city schools of Detroit and Chicago, also from the suburbs of Winnetka, Illinois and Scotch Plains, New Jersey. They represented every part of the country and nearly every sub-group of the nation's population.

Vermont Meeting--August 16 and 17, 1990

Standard setting is a judgmental process. This meeting in Essex Junction, Vermont, provided background and a framework for the 63 panel members to share their judgments. The meeting included the following:

- Judges received training about the process.
- Panelists met in four small heterogeneous groups at each grade-level--4, 8, and 12. The groups were given the actual test items from the 1990 math assessment. Each judge was

asked to make a first round of ratings, indicating what proportion of students at each achievement level ought to answer each particular question correctly. These ratings were aggregated and averaged to produce a recommended average percent correct for each achievement level at each grade.

o The groups were given information on how students actually performed on each question during the 1990 testing. Each judge then did a second round of ratings. During this round little or no discussion took place. Having performance information caused very little overall change in the ratings.

o The judges completed a third round of ratings. This time the judges participated in a discussion about their first two rounds of ratings.

o The results of the third round of ratings were shared with the total group of judges from all three grades.

Post-Vermont Meeting

o Revisions were made in some procedures based upon discussions with the technical advisory committee on standard setting. Two concerns were given special attention:

(1) Making sure judges had a clear understanding of the Board's general definitions of basic, proficient, and advanced.

(2) Ensuring that judges based their ratings on the difficulty of test items and their importance in showing mastery rather than on whether an item or item format was "appropriate" for NAEP.

o Analyses were conducted of the item ratings to be used at next panel meeting.

First Washington Meeting--September 29-30, 1990

o Thirty eight of the 63 judges reconvened in Washington.

o The entire group held a two hour discussion to clarify the definitions of basic, proficient, and advanced.

o Judges completed a fourth round of rating individual questions.

o Judges met by individual grade-level and in groups that included panelists from all three grades to discuss consistency and coherence of recommended levels.

o Judges made a fifth round of ratings giving the overall percent correct that ought to be required to reach each achievement level for their grade.

o Judges completed an evaluation form expressing their confidence levels in their final ratings.

Second Washington Meeting--November 12 and 13, 1990

o A smaller group of 11 judges wrote descriptions of the three achievement levels for each grade based on analyses of individual item ratings and average percents correct derived from judges' rating forms at earlier meetings. Sample items were selected to illustrate each proposed level.

o Final recommendations were sent to all panel members for approval.

NAGB Board Meeting in Atlanta--November 16 and 17, 1990

o The recommended achievement levels were presented to NAGB at its meeting in Atlanta. The Board also heard comments from the project's lead consultant, Professor Ronald K. Hambleton, of the University of Massachusetts, and from the lead evaluator, Professor Daniel Stufflebeam, of Western Michigan University. The evaluators recommended moving forward to completion but cautioned NAGB to proceed slowly enough to allow extensive public input.

o First public hearing was held on November 26.

Projected Next Steps

o Second public hearing will be held in Washington on January 8, 1991.

o In mid-January an interim report by team of independent evaluators will be made available to NAGB and to the public. Evaluation group is comprised of Stufflebeam (Western Michigan), Richard Jaeger (University of North Carolina), and Michael Scriven (Stanford).

o On January 22, 1991 the NAGB Executive Committee and Achievement Levels Committee will review public comment --from two hearings and written comments--and the interim evaluation report.

o In late January or early February of 1991 the NAGB expects to make a final decision regarding the mathematics achievement levels. If the levels are approved, the Board plans to issue a report that displays data from the 1990 math assessment in terms of these achievement levels. The report would be in its own booklet, separate from National Center for Education Statistics (NCES) reports on national and state NAEP results.

o By February 15, 1991 each participating state may decide whether to have its own 1990 NAEP eighth grade mathematics data prepared for use by the Governing Board in the NAGB report on achievement levels. After information is prepared, state superintendents will have a chance to approve actual data for release.

o The national and state data based on achievement levels could possibly be released as early as June 1991 or at the latest on September 24, 1991, the second anniversary of the Charlottesville Education Summit.

MAJORITY MEMBERS

DALE E. KILDEE, MICHIGAN, CHAIRMAN

GEORGE MILLER, CALIFORNIA
PAT WILLIAMS, MONTANA
MATTHEW G. MARTINEZ, CALIFORNIA
CARL C. PERKINS, KENTUCKY
CHARLES A. HAYES, ILLINOIS
THOMAS C. SAWYER, OHIO
MAJOR R. OWENS, NEW YORK
NITA M. LOWEY, NEW YORK
JOLENE UNSOELD, WASHINGTON
WILLIAM J. JEFFERSON, LOUISIANA
JOHN F. REED, RHODE ISLAND
TIM ROEMER, INDIANA
CRAIG A. WASHINGTON, TEXAS
PATSY T. MINK, HAWAII
JAIME B. FUSTER, PUERTO RICO
WILLIAM D. FORD, MICHIGAN, EX OFFICIO

(202) 225-4368

SUSAN A. WILHELM, STAFF DIRECTOR

COMMITTEE ON EDUCATION AND LABOR
U.S. HOUSE OF REPRESENTATIVES
320 CANNON HOUSE OFFICE BUILDING
WASHINGTON, DC 20515

SUBCOMMITTEE ON ELEMENTARY, SECONDARY
AND VOCATIONAL EDUCATION

MINORITY MEMBERS

WILLIAM F. GOODLING, PENNSYLVANIA
SCOTT L. KLUG, WISCONSIN
THOMAS E. PETRI, WISCONSIN
MARGE ROUKEMA, NEW JERSEY
STEVE GUNDERSON, WISCONSIN
STEVE BARTLETT, TEXAS
PAUL B. HENRY, MICHIGAN
SUSAN MOLINARI, NEW YORK
JOHN A. BOEHNER, OHIO

March 15, 1991

Daniel M. Koretz
Senior Social Scientist
The Rand Corporation
2100 M Street, N.W.
Washington, D.C. 20037-1270

Dear Dr. Koretz:

Thank you very much for the excellent testimony you presented at the March 13, 1991 hearing of the Subcommittee on Elementary, Secondary, and Vocational Education. The Subcommittee received a wealth of valuable information which will greatly assist our efforts to review the National Assessment of Education Progress.

Enclosed please find the additional questions which I indicated I would be asking you to respond to for the record. Due to the great interest in the issue, I would appreciate your assistance in returning the answered questions by March 22, 1991 so the hearing record can be printed quickly.

Again, thank you for your valuable testimony and your assistance in expediting the return of your answers to the additional questions.

Sincerely,

Dale E. Kildee
Chairman

dt
enclosure

QUESTIONS FOR DR. KORETZ

QUESTIONS

1) IN YOUR OPINION, IS THE USE OF NAEP AS A "THERMOMETER" TO MEASURE THE STATUS OF U.S. EDUCATION COMPATIBLE WITH AN EXPANDED USE OF NAEP AS A HIGH STAKES NATIONAL EXAMINATION COMPARING PERFORMANCE OF INDIVIDUAL SCHOOLS OR INDIVIDUAL STUDENTS?

2) IN YOUR TESTIMONY, YOU STATED THAT USING NAEP FOR ACCOUNTABILITY MIGHT UNDERMINE IT, BECAUSE PEOPLE WOULD STRIVE TO RAISE SCORES. BUT ISN'T THAT EXACTLY WHAT WE SHOULD WANT: FOR STUDENTS TO STRIVE TO ACHIEVE MORE?

3) IN YOUR TESTIMONY, YOU STATED THAT NAEP CANNOT PROVIDE RELIABLE INFORMATION ABOUT THE EFFECTIVENESS OF STATE OR LOCAL EDUCATIONAL PROGRAMS. MANY PEOPLE ARGUE OTHERWISE. PLEASE EXPLAIN IN MORE DETAIL WHY YOU BELIEVE THAT NAEP CANNOT SERVE THIS FUNCTION.

4) WHAT ARE THE MOST ACCURATE COST ESTIMATES OF A COMPREHENSIVE NATIONAL TEST?

THE RAND CORPORATION

March 20, 1991

Honorable Dale E. Kildee
Chairman
Subcommittee on Elementary, Secondary, and Vocational Education
Committee on Education and Labor
U.S. House of Representatives
Washington, D. C. 20515

Dear Mr. Chairman:

In response to your request, I am writing to answer the supplementary questions you sent me to follow up on the March 13, 1991 hearing on national assessment issues. I am pleased to have had the chance to be of assistance to you and the Subcommittee as you work on these issues.

Question

> In your opinion, is the use of NAEP as a "thermometer" to measure the status of U.S. education compatible with an expanded use of NAEP as a high-stakes national examination comparing the performance of individual schools or individual students?

Answer

No; these uses are incompatible. Evidence continues to accumulate showing that a single test cannot simultaneously serve both as a high-stakes tool of accountability and a trustworthy indicator--a "thermometer"--for gauging what our students know.

A primary reason why these two functions are incompatible is that when tests are used for accountability, educators begin to focus attention on the specific content of the test. The content of any single test, however, is just a small sample of what we want students to know. We infer from students' performance on that small sample what they know about the much larger "domains" of achievement that are really of concern to the

202 296 5000

2100 M Street NW
Washington DC 20037-1270

Main Office 213 393 0411
1700 Main Street PO Box 2138
Santa Monica CA 90407-2138

public and to policymakers. We are not concerned with the performance of high school seniors on the roughly 240 specific items that constitute the NAEP science test; we are concerned with what that implies about students' mastery of biology, chemistry, physics, and so on. When educators begin to focus on the specific content of the test, that generalization from the small sample of test items to the big domain of achievement is undermined. Scores become inflated.

In my view, extending the NAEP to provide comparisons of individual schools could undermine the value of NAEP as a thermometer even if no explicit sanctions were attached to it. Because of the current widespread use of tests as indicators of school quality, we have seen many instances over the past decade where simply publicizing scores was sufficient to convert a test into a "high stakes" examination. Providing comparisons of individual students would constitute even more of a risk, but it would also require a wholesale alteration of the test, because the NAEP's current design does not provide reliable estimates for individual students.

Question

> In your testimony, you stated that using NAEP for accountability might undermine it, because people would strive to raise scores. But isn't that exactly what we should want: for students to strive to achieve more?

Answer

I certainly agree that we need to find ways to encourage students to strive to achieve more. The question, however, is whether using NAEP for accountability is a sensible way to accomplish that goal.

One reason not to use NAEP in this manner is the problem noted in the previous answer: doing so will undermine the value of the NAEP as a thermometer of the achievement of our nation's youth. That would be a high price to pay, because we currently have no reasonable substitute for the NAEP as a thermometer. We have no other assessment that tests large, nationally representative samples of students at frequent intervals in many subjects. If we want to know whether or not the nation's educational achievement is improving, we need to preserve the NAEP as a thermometer.

A second reason not to use the NAEP for accountability is that test-based accountability has not shown itself to be an optimal way of encouraging students and teachers to learn more. Rather, it encourages them to raise scores, which sometimes does and sometimes does not entail commensurate increases in real achievement. Moreover, it sometimes has pernicious side-effects, such as narrowing of instruction and excessive drill on material closely related to the test.

Question

> In your testimony, you stated that NAEP cannot provide reliable information about the effectiveness of state or local educational systems. Many people argue otherwise. Please explain in more detail why you believe that NAEP cannot serve this function.

Answer

The specific reasons are numerous, but the general reason is this: many factors will cause states to differ in terms of test scores, and the NAEP is simply not designed to sort them out.

This becomes clearer if one starts with an example outside of education. Recently, the press reported that the mortality rate for adult males is higher in Harlem than in Bangladesh. Would one infer from that finding that hospital care in Harlem is inferior to that in Bangladesh, or that one is better off to become seriously ill in Bangladesh than in Harlem? Probably not. There are a great many factors that enter into adult male mortality rates, including, to name only a few, drug and alcohol use, motor vehicle use, access to handguns, access to primary health care, and the quality of hospital care. The difference in mortality rates has many plausible explanations other than the quality of hospital care.

Similarly, many diverse factors enter into differences in test scores, including, to name a few, parental education and aspirations, poverty, nutritional and health status, family composition, community values, and educational quality. The NAEP is simply not designed to disentangle these influences. It includes only limited and weak background information--for example, no information at all on individual income or poverty, only weak information about school-level poverty, weak information about parental education, and extremely limited information about community characteristics. One way to offset partially some of these weaknesses would be to use *longitudinal* data that permit one to track students' growth over time, but NAEP lacks that as well. Finally, NAEP collects only limited information about educational policy and practices. As a result, even when differences among states do reflect educational factors, we will often not have any clear idea *which* educational factors are to blame.

Ignoring these limitations and treating NAEP results--for example, the results of the Trial State Assessments--as indications of differences in educational effectiveness runs the risk of emulating ineffective policies and eliminating effective ones.

Question

> What are the most accurate cost estimates of a comprehensive national test?

Answer

This question has no simple answer; it depends on the test's characteristics. Performance assessment, for example, will cost more than conventional multiple-choice tests because of the added costs of task development and scoring. Additional test security also adds costs, by increasing logistical costs and often by requiring the development of additional test materials. Truly standardizing administration, to avoid bias from incorrect test administration, further increases costs; that is one factor that makes the current national NAEP more expensive than off-the-shelf commercial tests administered by teachers.

Many of the important costs of an additional comprehensive national test, however, are non-financial. As it is, our nation has an unusually short school year and fills an appreciable part of it with test-related activities. The time required for testing itself is often a small part of the burden, because schools often devote substantial time to test preparation. As we continue to add additional layers of testing, we need to ask about the time and local resources they require, what activities are being supplanted, and whether schooling will really be improved as a result.

Respectfully,

Daniel Koretz
Senior Social Scientist

MAJORITY MEMBERS
DALE E. KILDEE, MICHIGAN, CHAIRMAN
GEORGE MILLER, CALIFORNIA
PAT WILLIAMS, MONTANA
MATTHEW G. MARTINEZ, CALIFORNIA
CARL C. PERKINS, KENTUCKY
CHARLES A. HAYES, ILLINOIS
THOMAS C. SAWYER, OHIO
MAJOR R. OWENS, NEW YORK
NITA M. LOWEY, NEW YORK
JOLENE UNSOELD, WASHINGTON
WILLIAM J. JEFFERSON, LOUISIANA
JOHN F. REED, RHODE ISLAND
TIM ROEMER, INDIANA
CRAIG A. WASHINGTON, TEXAS
PATSY T. MINK, HAWAII
JAIME B. FUSTER, PUERTO RICO
WILLIAM D. FORD, MICHIGAN, EX OFFICIO

(202) 225-4388

SUSAN A. WILHELM, STAFF DIRECTOR

COMMITTEE ON EDUCATION AND LABOR
U.S. HOUSE OF REPRESENTATIVES
320 CANNON HOUSE OFFICE BUILDING
WASHINGTON, DC 20515

SUBCOMMITTEE ON ELEMENTARY, SECONDARY,
AND VOCATIONAL EDUCATION

MINORITY MEMBERS
WILLIAM F. GOODLING, PENNSYLVANIA
SCOTT L. KLUG, WISCONSIN
THOMAS E. PETRI, WISCONSIN
MARGE ROUKEMA, NEW JERSEY
STEVE GUNDERSON, WISCONSIN
STEVE BARTLETT, TEXAS
PAUL B. HENRY, MICHIGAN
SUSAN MOLINARI, NEW YORK
JOHN A. BOEHNER, OHIO

March 15, 1991

Robert L. Linn, Ph.D
Professor
University of Colorado
Campus Box 249
Boulder, CO 80309

Dear Dr. Linn:

Thank you very much for the excellent testimony you presented at the March 13, 1991 hearing of the Subcommittee on Elementary, Secondary, and Vocational Education. The Subcommittee received a wealth of valuable information which will greatly assist our efforts to review the National Assessment of Education Progress.

Enclosed please find the additional questions which I indicated I would be asking you to respond to for the record. Due to the great interest in the issue, I would appreciate your assistance in returning the answered questions by March 22, 1991 so the hearing record can be printed quickly.

Again, thank you for your valuable testimony and your assistance in expediting the return of your answers to the additional questions.

Sincerely,

Dale E. Kildee
Chairman

dt
enclosure

QUESTIONS FOR DR. LINN

QUESTIONS

1) CAN YOU DESCRIBE SOME OF THE REACTIONS YOU HAVE RECEIVED FROM STUDENTS, SCHOOL TEACHERS, ADMINISTRATORS, AND POLICYMAKERS DURING THE TRIAL STATE ASSESSMENTS?

2) DO WE RUN THE RISK OF OVERBURDENING OUR STUDENTS AND TEACHERS WITH TESTING? HOW MUCH TIME IS ALREADY CONSUMED BY TESTING IN THE AVERAGE STUDENT'S ACADEMIC YEAR? HOW MUCH MORE TIME MIGHT BE TAKEN UP WITH TESTING IF WE MOVED TO A SYSTEM OF NATIONAL ASSESSMENT OF ALL STUDENTS?

3) IN YOUR OPINION, IS THE USE OF NAEP AS A "THERMOMETER" TO MEASURE THE STATUS OF U.S. EDUCATION COMPATIBLE WITH AN EXPANDED USE OF NAEP AS A HIGH STAKES NATIONAL EXAMINATION COMPARING PERFORMANCE OF INDIVIDUAL SCHOOLS OR INDIVIDUAL STUDENTS?

4) HOW DOES LACK OF MOTIVATION AFFECT THE ACHIEVEMENT LEVEL OF STUDENTS, PARTICULARLY HIGH SCHOOL SENIORS? CAN YOU DESCRIBE WHAT NAEP IS DOING TO CONFRONT THIS IMPORTANT ISSUE IN THE STATE TRIAL ASSESSMENTS?

Response to questions posed by

Subcommittee on Elementary, Secondary,
and Vocational Education
Committee on Education and Labor
U.S. House of Representatives

Prepared by

Robert L. Linn
Professor, University of Colorado at Boulder
and Co-Director, Center for Research on Evaluation,
Standards, and Student Testing

March 21, 1991

1) CAN YOU DESCRIBE SOME OF THE REACTIONS YOU HAVE RECEIVED FROM STUDENTS, SCHOOL TEACHERS, ADMINISTRATORS, AND POLICYMAKERS DURING THE TRIAL STATE ASSESSMENT?

Only preliminary information is available at this time based on observations of some administrations, analyses of questionnaire responses from local administrators of the assessment, and a survey of state testing directors. Additional survey data will be collected from policymakers following release of the results in June and a more complete report will be included in the October, 1991 report of the National Academy of Education Panel evaluating the Trial State Assessment. The ways in which the data are received and used by policymakers, the press, and the public will provide one of the acid tests of the usefulness of the Trial State Assessment. Based on preliminary results, however, it can be said that most reports indicated that the administration of Trial State Assessment went well. Most state testing directors indicated that they thought that the Trial State Assessment results would be of value to their states.

School and student participation rates provide an indirect indication of reactions to the Trial State Assessment. Both school and student participation rates were good. Approximately 94% of the sampled schools agreed to participate and substitute schools were successfully obtained for some of the refusals. Approximately 94% of the students randomly selected and asked to participate agreed to take part in the assessment. As is now done with the national NAEP results and other surveys, statistical adjustments will be made for non-response. By way of comparison, it might be noted that the corresponding grade 8 participation rates for the 1988 national reading assessment were 86.6% for schools (prior to replacement) and 87.8% for students (based on the

1990 NAEP report "Learning to Read in Our Nation's Schools", Table A.1).

2) DO WE RUN THE RISK OF OVERBURDENING OUR STUDENTS AND TEACHERS WITH TESTING? HOW MUCH TIME IS ALREADY CONSUMED BY TESTING IN THE AVERAGE STUDENT'S ACADEMIC YEAR? HOW MUCH MORE TIME MIGHT BE TAKEN UP WITH TESTING IF WE MOVED TO A SYSTEM OF NATIONAL ASSESSMENT OF ALL STUDENTS?

Yes, we run the risk of overburdening our students and teachers with testing! Students in many school districts are required to take a standardized achievement test battery each year. Depending on how comprehensive the test battery is, anywhere from 2 to 5 hours of actual testing time may be required. The National Commission on Testing and Public Policy recently estimated that "each year elementary and secondary students take 127 million separate tests as part of standardized test batteries mandated by states and districts. (A test battery may include separate tests that measure such things as reading, mathematics, science, and study skills). At some grade levels a student may take as many as seven to 12 such tests a year." In addition, students may be required to take district or state tests as part of minimum competency testing requirements and state or district assessments.

The time spent in taking the actual tests, however, is but the tip of the iceberg when it comes to time devoted to externally mandated tests. Research conducted by my colleague, Lorrie Shepard, through the Center for Research on Evaluation, Standards, and Student Testing has shown that many elementary school teachers devote the equivalent of several weeks of class time teaching students test-taking strategies. In addition, in situations where accountability tests have a great deal of importance, teachers may teach content throughout the school year using worksheets that closely resemble standardized test formats.

How much more time might be taken up with testing if we moved to a National Test of all students would depend on the nature of that program and whether it was an addition to, rather than a replacement for, some existing testing. In its present form, NAEP requires only about one hour per student assessed. Because different samples of students are administered different samples of assessment exercises, this is adequate time for the purpose of monitoring achievement trends for the nation or a state. Substantially more time, say 3 or 4 hours, per student would be needed, however, in order to obtain reliable scores for individuals. If this were not only an add-on to current testing, but another test for which teachers felt compelled to spend time providing students with special test preparation, the expansion of National Assessment to all students would add significantly to an already excessive amount of instructional time devoted to externally mandated testing and preparation for those tests.

3) IN YOUR OPINION, IS THE USE OF NAEP AS A "THERMOMETER" TO MEASURE THE STATUS OF U.S. EDUCATION COMPATIBLE WITH AN EXPANDED USE OF NAEP AS A HIGH STAKES NATIONAL EXAMINATION COMPARING PERFORMANCE OF INDIVIDUAL SCHOOLS OR INDIVIDUAL STUDENTS?

No! There is substantial experience with testing programs that have been designed to serve multiple purposes. That experience indicates that a single testing program cannot simultaneously serve well the distinct purposes of monitoring trends in achievement, school building accountability, and the certification of individual student achievement. The accountability or certification functions result in a narrow focusing of attention on the test. As a consequence, generalizations about achievement defined more broadly than just the items on a specific test are undermined. That is, the validity of conclusions about educational progress that can be drawn from an independent indicator system such as NAEP would be threatened by its expanded role as a high-stakes national examination system.

In addition to a narrow focus on the content of the test, high-stakes testing programs are also subject to other pressures that limit their value as an independent monitor. As the stakes go up, there are increasing pressures to narrow the content of the test. Law suits associated with high-stakes testing programs have a chilling effect that reduces tests to a lowest common denominator. Specifically, they must be limited to what certainly has been taught rather than to what ideally should be taught. If this were to happen to NAEP, we would no longer be able to monitor progress in important higher-order thinking skills and the ability of students to solve novel problems. Yet it is precisely such accomplishments that are vital to the future of the nation.

4) HOW DOES LACK OF MOTIVATION AFFECT THE ACHIEVEMENT LEVEL OF STUDENTS, PARTICULARLY HIGH SCHOOL SENIORS? CAN YOU DESCRIBE WHAT NAEP IS DOING TO CONFRONT THIS IMPORTANT ISSUE IN THE STATE TRIAL ASSESSMENTS?

Assessments of student achievement depend on the cooperation of students, not only in agreeing to participate, but in the effort that they devote to the assessment. Questions have been raised about the motivation of students to perform their best on an assessment that does not have consequences for the individual participants. NCES has created a Technical Review Panel through a contract with the Center for Research on Evaluation, Standards, and Student

Testing and the RAND Corporation to conduct validity studies for NAEP. It is anticipated that studies of student motivation will be one of the priority areas for investigation. That work is yet to be done, however.

Student participation rates by grade provide some indirect information that is relevant to this issue of likely differential motivation for high school seniors in comparison to students at grades 4 or 8. In the 1988 reading assessment, for example, the participation rates at grades 4, 8, and 12 were 92.8%, 87.8%, and 78.5%, respectively (based on the 1990 NAEP report "Learning to Read in Our Nation's Schools", Table A.1). Clearly, high school seniors are much more likely to refuse to participate in NAEP than are their counterparts at earlier grades. The lower level of motivation of seniors to participate in the assessment might also be reflected in the effort of some of the students who do participate. This difference in student willingness to participate is not surprising to anyone familiar with high school seniors. Nor would it be surprising to find that the fraction of students who participate in the assessment in a half-hearted manner increased from 4th to 8th to 12th grades.

In any event, there are sufficient differences between high school seniors and students in the 4th or 8th grades that it would be unwise to make generalizations about the likely success of a Trial State Assessment for 12th grade students from results obtained for 8th grade students in 1990 or 4th and 8th grade students in 1992. Thus, if Congress is interested in expanding state-level use of NAEP to include the 12th grade, I would encourage consideration of starting with a trial in 1994 along the lines of the trials authorized at the earlier grades for 1990 and 1992.

Respectively submitted,

JOHN F. AKERS
STATEMENT TO THE HOUSE EDUCATION AND LABOR SUBCOMMITTEE ON ELEMENTARY, SECONDARY AND VOCATIONAL EDUCATION ON THE NATIONAL ASSESSMENT OF EDUCATION PROGRESS

Chairman Kildee, my name is John Akers, chairman and chief executive officer of International Business Machines Corporation. I am submitting testimony as chairman of the Education Task Force of The Business Roundtable (BRT). The BRT is an organization comprised of 212 chief executive officers of many of America's largest corporations. The BRT CEOs have committed their personal time and company resources to a decade-long effort to help improve the quality of the nation's K-12 education system.

The BRT has identified strategies we believe are essential to achieving the national education goals. I am submitting these strategies, "The Essential Components of a Successful Education System," for the record. We are working in all 50 states and the District of Columbia to address the essential components through legislation and regulation.

Assessment is a key point in our agenda. To provide assistance, we need detailed information about what students know and are able to do. And, to hold educators and policy makers accountable for student performance, we must have data to measure it at the building, district, state, and national levels. Thus, we applaud the Senate's interest in assessment and recommend that you allow the National Assessment of Education Progress (NAEP) to expand to measure student performance at the state, district, and building levels.

National testing has become a subject of significant debate. A consensus is emerging in favor of a national assessment system that would revolve around a common set of standards for student achievement. States, districts, and other entities would create exams, either jointly or independently, to measure whether students are meeting the common standards. Until such a system is developed, we have to rely on existing instruments.

NAEP provides a clear snapshot of what America's students know and are able to do. Educators and policy makers applaud its focus on writing and problem-solving skills.

NAEP is not perfect, but it is currently the best test of its kind. The National Education Goals Panel is considering using it as an interim measure. At their Winter Meeting in February, the National Governors' Association passed a resolution calling for the expansion of NAEP.

Congress should remove some of the restrictions on NAEP. NAEP currently publishes its findings about national student performance in the aggregate and breaks down student achievement by sex, race, region, and other categories. With the exception of the 1990 and 1992 pilot exams, federal law expressly prohibits NAEP from reporting student achievement by state.

We urge you to extend the state-by-state assessment to include all appropriate subject areas at all grade levels. We also support lifting the prohibition against using NAEP to measure district and building performance, so long as it can be done within the context of the present matrix sampling approach. The expansion of NAEP should be contingent upon a positive evaluation of the 1990 pilot test. Districts and buildings that participate in NAEP should be required to use the data to improve the performance of the most needy students. We do not support any use of NAEP that would allow for the comparison or evaluation of individual students.

Quality assessment instruments alone will not improve student achievement. But without stronger and richer assessment strategies, America's education system will not reach its potential.

We appreciate this opportunity to submit testimony and look forward to working with the Congress on this important issue.

March 14, 1991

FairTest
National Center for Fair & Open Testing

March 18, 1991

Rep. Dale Kildee, Chair
Subcommittee on Elementary and Secondary Education
320 Cannon House Office Building
Washington, DC 20515

Dear Rep. Kildee:

Thank you for having invited me to testify on the issue of a national test or examination system.

In addition to the testimony I have submitted on that issue, I request that the enclosed "Open Letter to Congress, Bush Administration, the Governors on NAGB and NAEP Expansion" be introduced into the record of the hearing on the National Assessment of Educational Progress of March 13, 1991.

The statement was endorsed by a long list of education and civil rights organizations and leading individuals. It responds to the expansion proposals set forward by the National Assessment Governing Board in December 1989 that were presented by Board member Herbert Walberg to the hearing on NAEP.

Thank you for keeping the record open and including this statement.

Yours,

Monty Neill

Monty Neill, Ed.D.
Associate Director

342 Broadway, Cambridge, Mass. 02139 (617) 864-4810 FAX (617) 497-2224

FairTest
National Center for Fair & Open Testing

OPEN LETTER TO CONGRESS, BUSH ADMINISTRATION, THE GOVERNORS ON NAGB AND NAEP EXPANSION

Over the past several months, the National Assessment Governing Board (NAGB) has taken several actions which, considered together, raise serious concern over the future direction of the National Assessment of Educational Progress (NAEP). As a group of education and civil rights organizations active in school reform issues, we are addressing our concerns to Congress, the Administration and the National Governors Association so that all responsible parties understand the nature of these problems and carefully monitor developments in NAGB and NAEP. It is important to note that we are not writing to oppose the national assessment, but to help ensure that it plays a constructive, not harmful, role in reforming our nation's educational systems.

The actions of the Governing Board, taken together, go far beyond the level of activity authorized in the National Assessment of Educational Progress Improvement Act adopted as part of the Hawkins-Stafford Elementary and Secondary Education Amendments of 1988. That Act (PL 100 - 297), which passed following lengthy discussion, authorized voluntary state-by-state comparisons of NAEP assessment results on a trial basis, and mandated an independent study of the validity and effects of the pilot programs.

Less than two years later, prior to completion of the trial comparisons and the studies, NAGB is proposing a major expansion of NAEP (see NAGB's paper "Positions on the Future of the National Assessment"). The proposal includes: 1) full participation by the states in state-by-state comparisons, to be paid for by the federal government; 2) testing and comparing local districts and even schools, which is currently prohibited by law; and 3) more frequent testing. Last month, NAGB adopted a process for setting "achievement levels" that students in grades four, eight and twelve ought to attain on NAEP tests (see NAGB paper, "Setting Appropriate Achievement Levels").

While each of these initiatives raises problems that require serious attention, we are particularly concerned about the *combination* of setting achievement levels and expanding NAEP. Our specific concerns and recommendations include:

1) The proposal to expand NAEP was adopted before completion of the Congressionally-mandated studies or the pilot state-by-state comparisons.

Expansion of NAEP will inevitably affect our nation's education. Congress correctly planned a cautious, step-by-step process to gauge the value and effects of state comparisons before mandating their continuation or expansion. This evaluation should be completed before any further steps are taken to expand NAEP.

2) NAGB is proposing expansion of NAEP before the national debate on educational goals is resolved.

So far, the Bush Administration and the Governors have agreed on broad national goals, but have yet to decide how to implement them. Logically, the Administration, the

Governors and Congress should all have roles in this debate as well as in determining the indicators used to measure progress toward the goals. But if measurement precedes goals clarification, the process of measuring becomes, by default, the process of defining. That would truly be putting the cart before the horse.

Deferring action on NAEP expansion until after the trial state comparisons and the legally required studies are completed will allow time for the national debate on attaining educational goals to reach resolution. Only then can NAEP play a constructive role in developing appropriate measurement tools and procedures.

3) It is reckless to consider lifting the ban on district-by-district or school-by-school comparisons without considering the consequences for curriculum and instruction.

No one yet knows the effects - and side-effects - even of state by state comparisons. Repeal of the ban on local comparisons requires much more information and public discussion. It should not be considered until after the results of all trial comparisons and the mandated studies have been fully analyzed and publicly discussed.

4) NAGB's achievement level setting process, when combined with comparisons, may create a *de facto* national curriculum.

The evidence is overwhelming that the more power attached to a test, the more control the test will have over curriculum and instruction. A national test with achievement goals and local comparisons will certainly become a powerful, perhaps controlling, influence on the curriculum.

The education goals enunciated by the Administration and the Governors do not attempt to mandate a national curriculum. In fact, there is widespread agreement that curriculum and instruction should not be determined from Washington. States and communities need flexibility in determining how to attain the broad goals. Yet NAGB's expansion proposals could preclude state and local initiatives.

5) NAGB's achievement level setting procedures for its tests are not appropriate for determining national achievement goals.

The process chosen by NAGB to set achievement levels on its tests relies on selecting items from existing NAEP exams that, in the view of committees of experts, should be answered correctly by students who have attained the levels of "basic," "proficient" or "advanced." This is not an appropriate method for determining national curricular goals and achievement levels because it allows one test to define the content area and what students should be able to do in that area. Such decisions should be made prior to and independently of any test. After curricular goals have been decided at the various levels, then assessments appropriate to the curriculum can be constructed and achievement levels set.

Moreover, as the recently-released report of the National Commission on Testing and Public Policy explains, the procedure of relying on committees of experts to examine items is flawed even for the purpose of setting cut-off scores on tests. NAGB thus expects a limited technical procedure to be adequate for shaping a national curriculum.

6) By setting achievement goals based on what are predominantly multiple-choice tests, NAGB runs the risk of defining national educational goals in terms dictated by these narrow instruments.

In potentially shaping curriculum and instruction, NAEP tests will affect both content and methods of teaching. Multiple-choice testing necessarily focuses on factual recall and simple comparisons and observations. It does not lend itself to revealing whether students know how to do something - to write a persuasive essay, research an historical event, or grasp the meaning of a scientific development.

The narrowness of these instruments has been recognized by the Governors, among many others, and has led to widespread efforts to develop and implement other means of assessment. If multiple choice testing continues to predominate, NAEP will provide a continual obstacle to teaching and assessing the important things students need to learn how to do. It will help perpetuate a reduced definition of the content to be studied and an entirely incorrect view of how students learn.

7) NAGB proposes to vastly increase the amount of its testing to include "at least three subjects each year."

The current NAEP authorization establishes a two-year testing cycle and a minimum frequency for testing various subjects. Only math and reading are to be tested every two years; other subjects are scheduled at four- or six-year intervals. Though its futures paper deferred discussion of the "exact configuration" of the new testing cycles, NAGB called for "testing at least three subjects each year," at least six tests every two years. NAGB's claims this acceleration is necessary "to provide timely and sufficient data" and to "replace the Education Department's annual 'wall chart' which relies on SAT and ACT scores."

Again, major changes in NAEP such as expanding the extent and frequency of testing should not be undertaken prior to completion and analysis of the 1992 testing and the mandated studies. In fact, such expansion is not at all necessary. Because educational systems and achievement cannot change rapidly, yearly aggregated data will not provide meaningful information about important educational changes. Less frequent information should be quite sufficient.

While virtually everyone, including Secretary Cavazos, agrees on the inadequacy of the current "wall charts," the mere existence of the charts is an insufficient justification for vastly increasing a national testing program. To be sure, annual one-point changes in average SAT scores or two-tenths of a point changes on the ACT in the "wall charts" are meaningless. But substituting minute changes in NAEP scores would not be an improvement. It could, however, produce public frustration and thereby jeopardize public support for educational reform. Maintaining NAEP's current, authorized schedule will provide as much useful information at less cost in dollars and, ultimately, in public credibility.

8) NAGB is moving too slowly in revising NAEP exams to rely less on multiple-choice questions and to develop other means of assessment which better measure the full range of knowledge and skills.

While NAGB claims that about 20% of this year's NAEP math items were open-ended, Paul LeMahieu, Pittsburgh's Director of Testing, informed the National Association of Test Directors that less than 5% were really open-ended items. The rest were multiple-choice questions with the answer options deleted. Like multiple-choice items, such questions are not very useful in measuring student abilities to use math to solve real-world problems.

Instead of expanding the use of outdated, multiple-choice tests, NAEP should become a leader in the national effort to develop improved forms of assessment that provide more information and do not endanger but rather enrich the curriculum. NAEP should work with the states, a number of which already have performance-based assessment projects under development, to produce and evaluate such assessments.

9) NAEP expansion will absorb an ever larger share of federal research and information dollars, but the results may not be worth the money.

The NAEP Improvement Act authorized $9,500,000 for fiscal year 1989 for NAEP. For FY 1990, NAEP received $17,084,000. Even with this increased amount, the Education Department deferred the NAEP validity study, a national assessment of adult illiteracy and work on the National Education Longitudinal Study. For FY 1991, NAGB has requested $18,866,000, an increase of more than 10% over FY90 and nearly double the authorization for FY89. NAGB receives up to 10% of NAEP funds for administrative purposes and reportedly seeks to receive up to 15%. Estimates of the cost of NAEP if expanded are $100 million annually, a more than five-fold increase over current expenditures and an amount two and one-half times the funding for the National Center for Educational Statistics (NCES).

Will the results be worth the additional money? Yearly testing will not increase anyone's knowledge of the effects of educational reform efforts. Further state and local comparisons may not tell us more than we already know about how well the states and localities perform on standardized tests. In a period of continuing fiscal restraint, money used for more extensive testing could be better used to improve the *quality* of NAEP assessments or for other needed research rather than for redundant and potentially dangerous increases in testing.

10) The relationship among NAEP, NAGB and NCES must be clarified.

The current debates over the future of NAEP have raised questions about the appropriateness of an independent body wielding the power that NAGB could assert over our nation's education. A key issue is whether such a body is adequately accountable to Congress, the Administration and the public.

Since accountability is, in part, asserted by control over funding, NAGB's budget should be separated from NAEP's. So long as NAGB obtains a percentage of a (potentially rapidly-expanding) NAEP budget, there is no way for elected officials to adequately exert oversight. The role of NAGB in relation to NCES, the Department of Education or any other bodies created to oversee progress toward national goals should be carefully considered by the appropriate House and Senate committees and the Administration before NAEP is expanded.

In sum, NAGB's plans to rapidly expand NAEP without adequate consideration of the effects of the expansion or the proper role of assessment in educational reform are dangerous. Neither Congress nor the administration should allow them to proceed without careful review and consideration. Similarly, the Governors should not support the use of NAEP for measuring progress toward national goals without first clarifying the goals and the role of assessment in achieving them and then determining the details of measurement. Specifically:

- NAEP should not be expanded to allow more frequent or extensive testing or more detailed comparisons at least until completion of the trial assessments of 1990 and 1992 and the independent evaluation mandated in the Act. Then, Congress, the Administration and the Governors must weigh carefully the potentially harmful effects of more extensive testing and comparisons and ascertain that the dangers do not outweigh any possible benefits. In any event, expansion of NAEP must be subsequent and subordinate to the establishment of national goals and not allowed to dictate a national curriculum.

- NAEP should be directed to spend a significant portion of its budget on developing and piloting performance-based assessments (including tests and portfolios). Such research and development should be planned carefully to coordinate with state projects such as those underway in California, Connecticut and Vermont, to develop performance-based assessments, as well as projects undertaken by local education authorities or other governmental or private bodies.

- Congress and the Administration should consider separating NAGB funding from NAEP funding and carefully consider the future role of NAGB in relation to other agencies and bodies.

We appreciate your attention to these most important issues and look forward to working with you in the effort to achieve genuine and lasting reforms in the quality of public education.

Please feel free to call any of us if you have any questions or need further information.

LIST OF SIGNERS

Ronald J. Abate, College of Education, Cleveland State University*
Advocates for Children of New York, Norman Rollins, Executive Director
American Association of Colleges for Teacher Education, David G. Imig, Executive Director
American Association of School Administrators
American Reading Council, Julia Palmer, President
Angela N. Ancheta, Asian Law Alliance*
APPLE Corps., Inc., Alfred E. McWilliams, President, MaryAnne F. Gaunt, Executive Director
Association for Supervision and Curriculum Development, Gordon Cawelti, Executive Director
Bank Street College of Education, Joseph Shenker, President
Rims Barber, Mississippi Human Services Agenda
Leonard Beckum, Dean, College of Education, City College of New York
Association for Women in Science, Stephanie J. Bird, President
Gerald W. Bracey, Director of Research and Evaluation, Cherry Creek, Colorado
Diana Caballero, Puerto Rican/Latino Education Roundtable, New York City*
Center for Law and Social Policy, Alan W. Houseman, Director
Center for Women Policy Studies, Leslie R. Wolfe, Executive Director
Harold E. Dent, Vice-President, Psychological and Human Resources Consultants, Inc.*
Educational Law Center, Inc. Marilyn Morheuser, Executive Director
Pablo Eisenberg, President, Center for Community Change*
Federation of Organizations for Professional Women
Foxfire Fund, Inc. Eliot Wigginton, President
Fund for the Feminist Majority, Eleanor Smeal, President
Howard Gardner, Project Zero, Harvard University*
Genesee Valley Developmental Learning Group, New York
Leslie A. Hart, Brain Compatible Education Associates
La Donna Harris, Americans for Indian Opportunity*
Asa Hilliard, Professor of Education, Georgia State University*
Holt Associates/Growing Without Schooling

Institute for Learning and Teaching, Wayne Jennings, Director
International Reading Association
KEY-Kids, Education and You, (Beth Bradley, Jenny Coston, Leslie Floyd, Sue Long)
A. Gay Kingman, National Congress of American Indians*
Nancy K. Klein, College of Education, Cleveland State University*
Massachusetts Advocacy Center, Stephen R. Bing, Executive Director
Matsushita Foundation, Inc., Dr. Sophie Sa, Executive Director
Deborah Meier, Principal, Central Park East Secondary School, New York City
Sara E. Melendez, Vice-Provost, University of Bridgeport*
Susan Metz, Prospect Heights High School, Brooklyn, NY*
Mexican American Women's National Assocaition
National Association for the Advancement of Colored People, Beverly Cole, Education Director
National Association of Secondary School Principals
National Center for Fair & Open Testing (FairTest), Cinthia Schuman, Executive Director
National Coalition of Advocates for Students
National Coalition of Title I/Chapter I Parents, Robert Witherspoon, Director
National Council for the Social Studies
National Council of Teachers of English, Charles Suhor, Executive Director
National Education Association, Keith Geiger, President
National Indian Youth Council, Inc., Cheryl J. Mann, Executive Director
National Organization for Women--New York City Chapter
National Parent Teacher Association
National Women's Law Center
National Women's Political Caucus
Fred M. Newmann, National Center on Effective Secondary Schools--University of Wisconsin*
New York Public Interest Research Group
NOW Legal Defense and Education Fund, Helen Neuborae, Executive Director
Organization of Chinese American Women, Faith Lee Breen, Chair, Board of Directors
Vito Perrone, Harvard University--Graduate School of Education*
Project Equality, Inc., Kansas City, MO
Puerto Rican Legal Defense and Education Fund, Inc., Ruben Franco, President/General Counsel
Rochester (NY) Teachers Association (AFT), Adam Urbanski, President
Lori Rubenstein, Partnership for Democracy*
William V. Schipper, Executive Director, Natl. Assn. of State Directors of Special Education*
Donald H. Smith, Chairman, Dept. of Education, Baruch College, City University of New York*
Hilton Smith, Coordinator, Foxfire Teacher Outreach Programs*
Southern Association on Children Under Six, Cathy Grace, Executive Director
Southern Christian Leadership Conference, Joseph Lowery, President
Southern Regional Council, Inc, Steve Suitts, Executive Director
Judy I. Stahlman, Cleveland State University*
Adria Steinberg, Harvard Education Letter*
Gail E. Thomas, Texas A & M University*
United States Student Association, Julianne Marley, President
Dorothy J. Watson, President, Whole Language Umbrella*
Paul Weckstein, Center for Law and Education*
Grant Wiggins, Center on Learning, Assessment and School Structures (CLASS)
Arthur E. Wise, Rand Corporation*
Women's Research and Education Institute, Betty Parsons Dooley, Executive Director

***Organization for identification purposes only**

HEARING ON NATIONAL TESTING: PROS AND CONS

THURSDAY, MARCH 14, 1991

House of Representatives,
Subcommittee on Elementary, Secondary,
and Vocational Education,
Committee on Education and Labor,
Washington, DC.

The subcommittee met, pursuant to call, at 10:10 a.m., Room 2175, Rayburn House Office Building, Hon. Dale E. Kildee [Chairman] presiding.

Members present: Representatives Kildee, Hayes, Owens, Unsoeld, Roemer, Goodling, Klug, and Petri.

Staff present: Susan Wilhelm, staff director; Andy Hartman; Lynn Selmser; S. Jefferson McFarland; Damian Thorman; Jack Jennings; Katrina Kelley; June Harris; Carole Stringer; and Beth Buehlmann.

Chairman Kildee. The Subcommittee on Elementary, Secondary, and Vocational Education will come to order.

As a former schoolteacher, I usually start right on time, so I apologize for the delay this morning and appreciate the witnesses bearing with me.

The subcommittee meets today for the second in a series of hearings on educational testing. Yesterday, we heard from expert witnesses about the National Assessment of Educational Progress, the only federally mandated test of student achievement. Today's hearing focuses on the pros and cons of national testing generally.

There are several national testing or assessment trains which all appear to be approaching the station at the same time. The first is guided by the need to measure progress toward the national educational goals which the governors presented to us. A second is driven by a perception that a basic skills test could give business a better indication of student achievement. A third promotes national testing as a way to force educational reform.

Today's hearing will attempt to address some of the questions raised by the cargo contained in each of these trains. That analogy may be stretched a bit, but we will use it anyway. It is the subcommittee's intention to carefully review the current educational testing system and to consider each of the positive and possible negative consequences of a national test.

I look forward to hearing from the many knowledgeable and distinguished witnesses we have here today. We are fortunate to have you with us.

I call upon now my friend and the ranking Republican member of the full committee and the ranking Republican member of this subcommittee, Mr. Goodling from Pennsylvania.

Mr. GOODLING. Thank you very much, Mr. Chairman, and welcome, all of you here this morning.

We had an excellent session yesterday, and, as I indicated yesterday, one of my greatest concerns at the present time is that all of the concerns in relationship to choice, national testing, national curriculum, national you name it, maybe out of frustration, say, "I don't know what to do; things are bad, and I don't really know what to do, so let's just try all these kinds of things and see whether something won't happen that will be better," rather than, you know, trying to improve education.

I guess I would preface that statement by saying that I get pretty upset with the American attitude with respect to public education. The American attitude in relationship to Americans is always the same; we always put down ourselves. I talk to people all over the world, and they wonder why we are so critical of ourselves.

We tend to lump all of public education into one little barrel and say there is something wrong with it, radically wrong with it, and we must do this, that, and something else. I looked at the list of my leadership, for instance. All of them graduated from public schools. I looked at the list of those who serve here on the committee, and on my side eight out of 12 were public school graduates, and therefore I say they were successful before they came here, they have been successful since they have been here, and, yes, we have a lot of problems, but I sure don't want anyone to take the Wall Street Journal "National testing: No longer a foreign idea" as some kind of gospel, because it is about as misguided as anything I have read. Of course, it starts out by saying that national education goals seemed plenty radical when President Bush and the governors announced them just a year ago.

I didn't know that they were ever considered radical. I thought we always were striving toward national goals. But then when the article really gets out into left field is, by saying "San Diego's education system should navigate by the same stars as Atlanta's," or "that a well-schooled youngster in Alabama should possess the same knowledge and skills as one in Minnesota."

I mentioned yesterday that my State got off on the wrong foot. The former secretary of education in my State decided that we should test all students and then rank our schools in Pennsylvania according to the rests of those tests. Of course, I guess he was surprised to find out that a school district in western Pennsylvania was number one. Well, if it isn't, it sure should be, because the per pupil expenditure is probably three times that in all but about three school districts in the Congressional district I represent. All the parents are college graduates, many of whom are Ph.D.'s. All of the parents insist from day one, before the child opens his eyes or her eyes, that they are going to strive for excellence. They insist that their child strives for excellence. They also insist that the school strives for excellence.

So my hope is, as we go through this, quote, reform process, and particularly the testing process, that we are thinking primarily in terms of, "How do I help this child with the information from the test results that I now have before me?" I always try to convince my teachers that their purpose for testing was to see where they may have failed in presenting the material that they wanted to present and where they might find that they can help individual youngsters because of the results of those tests.

I will repeat the story I told yesterday also of my first experience as a guidance counselor. The superintendent mandated that I group all sixth graders that were moving into seventh grade. He didn't ask me whether I thought that was a good idea or not; he just said, "Do it," and I started it and discovered that one class from Stewartstown appeared to already have been grouped, and when I asked the superintendent if that was true, he said, "Oh, no; you just have to understand, Mrs. so-and-so teaches to the test from day one."

Now I don't have any problem teaching to the test if there are some specific goals that we want to reach and some knowledge that we think they should have, et cetera, et cetera. Any teacher would teach to any test that they are going to give. But I want to make sure that it is a tool to help the youngster, and I want to make very, very sure that we don't compare apples with oranges, which is what our secretary of education in Pennsylvania tried to do. He is no longer, I might add, the secretary of education. He embarrassed the governor, so he is no longer that secretary.

Having said all of that, I welcome you to the hearing this morning and look forward to your testimony.

Chairman **KILDEE**. Major Owens.

Mr. **OWENS**. Yes, Mr. Chairman. I ask unanimous consent to submit my written statement for the record and would just like to make a brief comment.

I want to thank you for holding this hearing on the question of a national test and yesterday's hearing on NAEP. This is an important issue and one which the Subcommittee on Select Education, which I chair, will also be examining during its consideration of the reauthorization of the Office of Educational Research and Improvement.

The Department of Education, for example, has proposed to rename the Office of Educational Research and Improvement to the Office of Assessments, Statistics, Research, and Improvement, and I fear that these kinds of efforts will subordinate the more aggressive efforts that are needed to put the latest research in the hands of practitioners and to use our substantial knowledge base to improve education rather than assess. The Department's new motto, it seems, is going to be, "Don't just do something; stand there and assess," and I think we have to take steps to guard against that.

Thank you.

[The prepared statement of Hon. Major R. Owens follows:]

PREPARED STATEMENT OF HON. MAJOR R. OWENS, A REPRESENTATIVE IN CONGRESS FROM THE STATE OF NEW YORK

I want to thank the Chairman for holding this hearing on the question of a national test and yesterday's hearing on NAEP. This is an important issue and one which the Subcommittee on Select Education, which I chair will also be examining during its consideration of the reauthorization of the Office of Educational Research and Improvement.

141

Though some of its proponents may be utterly sincere, I cannot help but view this new wave of interest in testing with great suspicion. I fear it is yet another effort by the White House and the right-wing to substitute symbolism for substance in education policy making and designed to divert our attention from the urgent need for a greater Federal investment in our children's education.

We have seen this, for example, in the President's proposed budget for OERI for fiscal year 1992. The department is proposing to rename the Office of Educational Research and Improvement to the Office of Assessment, Statistics, Research, and Improvement. Aggressive efforts to put the latest research in the hands of practitioners and to use our substantial knowledge base in this area to improve the education of our children would be abandoned in favor of the passive role of merely assessing the status of education in this country. The Department's new motto, it seems, is don't just do something, stand there—and assess.

As they sit in overcrowded, dilapidated classrooms, children in New York already spend an average of four weeks every school year taking various standardized tests. There is something ridiculous about the idea that what they most need now is to have another, still more exotic test shoved under their noses in the name of holding them accountable for their performance. Our preeminent concern at this moment should instead be with holding politicians and policy-makers at every level accountable for their consistent failure to give our children and our schools all of the resources they need to succeed.

Chairman **Kildee**. Thank you.

Mr. Petri.

Mr. **Petri**. Thank you, Mr. Chairman.

I think we are having a hearing on an important subject, and I am looking forward to the testimony of the panels that we have today and to learning, particularly with the point of view of doing something, if we do it, that will improve education rather than doing something just for the sake of doing something in the vague hope that it might make things better, when the reality might be that it might make things worse.

But let's proceed and learn and see if we can do something that is useful for the young people in our country.

Chairman **Kildee**. Thank you.

Mrs. Unsoeld.

Mrs. **Unsoeld**. Thank you, Mr. Chairman.

I, too, am looking forward to hearing the testimony, because this is probably a topic about which I know the least with which we are dealing. So welcome, panel, and the other experts.

Chairman **Kildee**. Mr. Hayes.

Mr. **Hayes**. Thank you, Mr. Chairman.

I do not have a prepared statement. I do want to commend you, as others have done, for the scheduling of this important hearing, and I want to commend the witnesses for giving of their time to be here with us.

I have some very deep-seated concerns about what I see as an effort to almost privatize the public educational system, which will diminish the opportunities for the economically disadvantaged and minorities from getting an opportunity for education. I am a little concerned that testing may be a step in that direction, and I want to be here and be a part of it, and I am just sorry I wasn't here at the beginning, but, as you well know, when you have five subcommittees you serve on and two main full committees, we have all found out it is impossible to be at two places at the same time, much less three. So I just had to be somewhere else, but I am here today and expect to get much out of this hearing.

Thank you very much.

Chairman **Kildee**. Thank you, Mr. Hayes.

Mr. Roemer.

Mr. **Roemer**. Thank you, Mr. Chairman.

I would just like to commend you and the ranking Member, Mr. Goodling, for holding this hearing. As a new Member, I am excited to be here working on education. This committee was my first choice. Education is the number one goal on the home front to address. Business people who are involved in educational efforts across the board in my community, come into my office to talk about improving education.

I would like to welcome the distinguished panelist, Mr. Brock, back to the Hill, and also the namesake, Governor Romer. I think we still have a couple of Roemers in the Democratic Party.

I assume you don't have any plans to announce anything today, Governor Romer.

Mr. **Romer**. I think the record should show that the majority of Romers have chosen to not switch but stay and fight.

Mr. **Roemer**. The bipartisan Roemer efforts to improve education.

I am excited about the panel's testimony and would just like to urge that we keep our focus on the long term and not the short-term political gain of coming up with an idea that looks good. We need to help more students who are coming out of school who want to enter the work force in a competitive economy.

Thank you, Mr. Chairman.

Chairman **Kildee**. Thank you very much.

Now we will go to our first panel, consisting of the Honorable Roy Romer, chair of the National Education Goals Panel, Governor of Colorado and the Honorable William E. Brock, well known here on the Hill, a former member of the Senate, former Secretary of Labor, with whom I had the privilege of working—productively working, too—and chairman of the Secretary's Commission on the Achievement of Necessary Skills; and Gregory R. Anrig—Dr. Anrig—president of Educational Testing Service, Princeton, New Jersey—my three children who are now in college back in Michigan have sweated through your test a couple of times—and one who brings great credentials to the panel here this morning, well known to myself: Trevor Sewell—Dr. Sewell—acting dean, College of Education, Temple University.

Governor Romer, you may begin.

Mr. **Romer**. May I use the chart? Is that permissible here, Mr. Chairman?

STATEMENTS OF HON. ROY ROMER, CHAIR, NATIONAL EDUCATION GOALS PANEL, DENVER, COLORADO; HON. WILLIAM E. BROCK, CHAIRMAN, SECRETARY'S COMMISSION ON THE ACHIEVEMENT OF NECESSARY SKILLS, WASHINGTON, DC; GREGORY R. ANRIG, PRESIDENT, EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY; AND TREVOR SEWELL, ACTING DEAN, COLLEGE OF EDUCATION, TEMPLE UNIVERSITY, PHILADELPHIA, PENNSYLVANIA

Mr. **Romer**. I wanted to use this chart. I know we need to be very strict in our time limits, and I will be.

First, let me give you the power of this issue. As chairman of the National Goals Panel, 50 governors have agreed to hold themselves accountable each year for 10 years on the national goals. It is this panel's obligation to frame that report card and have it published between now and September. So our obligation is to take each of those goals and set specific questions that 50 States will answer, and that is real power, because the way you frame the question will set the agenda for those governors over a 10-year period. So the power of this panel is very, very large.

Now, we have both a short-term and a long-run assignment. Short term, by September 1 we need to have a report card, and we will give it, but it is obvious that we do not have the data, the indicators, available in this Nation, or even State by State, to give an accurate picture. Therefore, our second assignment is long term, to develop an assessment system for this Nation to where we can know where we are but, more important, where we can reform our system to reach where we ought to be.

Now, that, therefore, is on our agenda immediately, and I would like then to share with you very briefly some thoughts we have upon the issue of a national test. I will give you my one-line conclusion. I don't think we should have a national test; I think we should have a national assessment system, and I would like to give you the outline for that system.

But first, I found it is important to talk in language that people on the street can understand, and I would like to give a couple of illustrations that point out what testing is to me. I happen to be a pilot, and when you are getting a pilot's license, it may take you 46 hours to do it or it may take you 36 hours to do it, but whatever time it takes, there is a set of knowledge and skills, a body of knowledge and skills, you have to acquire.

I don't think any of you want to fly with an above-average pilot. You don't want a pilot who isn't at the top end of the bell curve. You want a pilot who has acquired certain knowledge and certain skills to master the craft. Therefore, I used to run a flight school. What do you do? You have an identifiable list of skills; you have a program of instruction that relates to that, and when you test, as you do, you are obviously using it for the purpose of assisting that student to know where he or she is in reference to where he or she ought to be.

So I think just to keep in mind, first of all, that testing in its most important function, I think, is instructional, it is diagnostic, and if we are thinking about creating a national assessment system, I think it is a secondary benefit to know where we are in terms of a sample as a Nation or a State. The primary benefit of a test is—can it be a part of the instructional tool? Can it be a part of the reform of the system?

Therefore, I just want us to begin to think about, all right, if we are going to be a pilot, there is a certain set of skills that we need to know, and it is obvious you can't test unless you know first what the skills are.

Therefore, let's turn that from piloting to math. When I go into a school, I often see the periodical chart of chemicals. I think sometimes we ought to take it down and put up this kind of a chart, and it is a chart on math, and every student in school can look at it and say, "Wait a minute; at the twelfth grade, I need to know and be able to do the following things: At the eighth grade, I need to know and be able to do these things; at the fourth grade, I need to be able to know and do those things."

You know, it is kind of like a medical CAT scan. I go to the doctor and take a CAT scan because there is an assumed level of health, and the CAT scan tells me where I am in reference to that and then says, through his prescription, "Here's how you become well."

I think educationally, a student in the fourth grade, the eighth grade, of the twelfth grade needs to have a clear understanding of what health is educationally. You need to know what the standards are you need to shoot for to be employable, to be a successful citizen, and that is our first job, to arrive nationally, through a consensual process, at standards of achievement. Then we will obviously find in math that there is a range of skills, and somebody may be on this end of the range and be a technician, and somebody may be on that end of the range and be an engineer, but if you are going through that system and you are at this level, you need to be forewarned that you are ill educationally, and you ought to take corrective measures, because you are not going to be a functional citizen.

So, again, I would like to move from piloting to math and say we need to think of testing and what it does most of all for the individual student and the parent of that student, because if we get that message across—namely, this is where you ought to be, and this is where you are, and we can make a change right now to get you on to the right track—that is the most benefit of testing.

Let me then turn to the question of, should we have a national test to do this, or should we find a better way to do it? And I will conclude with this. When I took this assignment as chair of this panel, I tried to identify what I thought was the most important barrier to accomplishing its mission, and I concluded that it was the ideological conflict in this country about local control for education and national goals, national standards; that is the conflict; and I said we ought to avoid that, we ought to avoid polarizing politically, and also substantively we ought to analyze what gives us the best product.

I looked at a number of alternatives, and I came to the conclusion that we ought to try what I would call a cluster method, and it is going to individual States, and let's assume that we are again dealing with fourth grade math; we are going to have to devise some new instrumentalities, because there are a lot of test instruments out there that, frankly, don't give us the right kind of assessment of the skills that we need to have.

Therefore, if we are going to create some more authentic assessment instruments of fourth grade math, including higher order thinking skills, there is no sense in our doing it individually, State by State. I think we can save some money by getting some States together and working together like an interstate compact.

Let's say you had six States that are brought together either by their educational philosophy, geography, or whatever, and they form a group—call them Group One—and they develop together an assessment instrument for fourth grade based upon their particular approach. Let's assume that there are four such groups in this country. There is a way that that can happen and we can tie them together through a national standards mechanism, and that is as follows: to create a set of national standards on fourth grade math.

Namely, we would arrive consensually at what it is a person ought to know and be able to do as a fourth grade math student. That, I think, can be done, and I think it needs to be done, and it ought not be done from the top down; it ought to be done consensually from the bottom up, and there are ways to do that. But we arrive at what it is that a person needs to know and be able to do as a fourth grade math student as a set of standards. So that then becomes a—let me just call it the standards box. We agree upon that consensually in a national assessment system.

But then we turn to these individual groups and say, "You create your own assessment instrument, but let's calibrate it to these standards so that when you are all done we know what a student who takes the test in one area would be as compared to all other areas," and that can be done. England has a similar system. They have about five or six different panels that you can go to and say, "I want to use your test." Those panels are informally calibrated to each other.

Very briefly, I think we can put together a cluster system in which we have the best of both worlds. First, we arrive at a national consensus on what the standard is we are reaching for; we leave it to the local district to decide instructionally how you get that done; and then we share this business of assessment so that we do not have one test nationally dictated from the top down, but that we go to the individual States and say, "Cluster; group together; use your own creativity; and then let's calibrate you through a standard box." And the way that calibration can be done is, you just simply create an anchor test which, for those who are experts in the field, will work quite well.

Now let me stop at this point and say, the reason that I throw out this suggestion, the panel has not yet accepted this as the way to go, but we are doing a lot of discussion about it. I think that it is a way this Nation can get together with this polarity of, how do you keep local control of education and local ingenuity and creativity and, at the same time, have some national direction as to what it is we ought to do in terms of levels of achievement and performance?

Thank you.

[The prepared statement of Hon. Roy Romer follows:]

STATEMENT OF

THE HONORABLE ROY ROMER

GOVERNOR OF COLORADO

BEFORE THE

ELEMENTARY, SECONDARY & VOCATIONAL EDUCATION SUBCOMMITTEE

COMMITTEE ON EDUCATION AND LABOR

U.S. HOUSE OF REPRESENTATIVES

A NATIONAL TEST: THE PROS AND CONS

MARCH 14, 1991

Mr. Chairman, members of the committee, thank you for the opportunity to testify before the Elementary, Secondary and Vocational Education Subcommittee of the U.S. House of Representatives Committee on Education and Labor. The topic we are discussing, the pros and cons of national education testing, is a vital one. I am pleased to take part in this debate with such distinguished colleagues as yourselves and my fellow panel members.

I have the privilege to serve as the co-lead Governor for education for the National Governors' Association and as chairman of the National Education Goals Panel. I want to make clear, however, that I am here today in my capacity as Governor of Colorado and am not speaking in any formal way for NGA or the Goals Panel. The ideas I will discuss have been, and will continue to be debated by the Governors and by the Goals Panel. They are creating interest and excitement, but neither the National Education Goals Panel nor the NGA have yet adopted a formal policy on the issue of national education testing. I want to begin by giving you a very brief background on the National Education Goals Panel. National Education Goals were established to provide a common framework and vision for education reform. As Governors, we wanted to be clear that we were committing ourselves to making education a priority in each of our states and were willing to be held accountable for this commitment.

The goals themselves have been criticized as too ambitious. In my judgment, they accurately identify what our reach must be to make the kind of educational progress that is required to secure our future. The goals are not gospel. They are not perfect. But they do provide a common vision, a common way to organize reform issues, and a common language.

Last July, the National Governors' Association adopted a policy establishing the National Education Goals Panel to oversee the development and implementation of a national education progress reporting system. Beginning in September 1991, the Panel will issue an annual progress report to the nation on our progress toward achieving the national education goals.

The Panel's work involves two tracks: one short-term and one long-term track. The short term agenda is to produce a credible report to be released in September 1991. In working toward this deadline, the Panel is aware that great harm can be done by asking the wrong questions or by trying to force existing data into new forms. Accordingly, the Panel is committed to a thoughtful and comprehensive approach over the long term, even if it means leaving some blanks in the first report.

Our long-term agenda involves encouraging the development of an assessment system that will not only measure our progress toward the goals, but that will help us restructure the education system to achieve the goals.

In an effort to take advantage of the expertise that exists in the country on these issues, the Panel established six working groups to give advice and policy options to the Panel about its work, both short-term and long-term. These groups are in the process of developing creative approaches for measuring our progress in each of the six goal areas. A list of those working groups is attached to this testimony. Their recommendations to the Panel will be released late this month.

With respect to goal #3, increasing student achievement, the Panel has identified three waves of action and effort that will need to be mobilized if we are to make substantial progress toward the goals by the year 2000:

- Defining specific standards of student achievement (what a student should know and be able to do);

- Designing an assessment system that will allow us to accurately and authentically measure our performance against these standards; and finally,

- Moving the education system to close the gaps between current levels of performance and the standards.

These tasks lie along a continuum from national to local effort. In the Panel's judgment, national goals and national standards for student performance are appropriate and necessary. At the other end of the continuum, we believe that the design and implementation of specific reform strategies must remain locally based in the best of our traditions. In the middle, the assessment challenge provides an opportunity for joint action. Let me discuss each of these in turn.

Standards. We need to begin with detailed goals and objectives for learning. These objectives must be uniformly high and must reflect what future citizens of our nation will need to know and be able to do. Given the diversity of reform strategies, an overall national educational standards framework is needed to provide the targets at which all would aim their efforts.

Discussion regarding the process by which these standards can be set has just begun. The following elements have been identified as critical to this process:

- There must be a key role for Governors and the states, which still have primary policymaking and funding responsibility for education.

- The process must be highly consensual, participatory and inclusive. Broad consultation with all segments of our population is the only way to build the commitment of all Americans to the hard job of educational change. In fact, the standard-setting process can itself become part of the renewal process if it represents a shared clarion call for higher expectations and higher standards for all students.

- To make sure the standards we set are competitive, we will need to use as a benchmark the best standards in the world.
- The process would set standards, not prescribe how the standards are to be achieved. Common standards need not and should not give rise to a national curriculum that dictates to teachers what and how they must teach or that limits the choices of school districts or states in selection of textbooks or instructional materials. The standards would be a description of the concepts, knowledge, and skills that students should master, not a recipe for achieving those standards.

Assessment. Tests should have the capacity to support reform and not just measure present performance. Tests also should have the capacity to measure the skills and abilities students will need to know.

I am concerned that many of the assessments we use now don't authentically measure higher order thinking skills. We rely on standardized tests that tend to drive curriculum and instruction toward rote memorization and the discovery of a single "right" answer. I am also concerned that the tests on which we rely to provide national and state samples of performance don't give the kind of feedback individual students need to help them and their parents judge the adequacy of their performance against national standards. These tests are useful and I support their continuation, but for assessment to stimulate better performance by students, the test ultimately has to be important to each student.

The notion of testing all students rather than a sample of students raises legitimate concerns about local control of education. There is a fundamental tension between the compelling need for national goals and standards and the tradition and strength of local control. We know that the most effective schools are those that are managed at the school level. We draw strength from the diversity of approaches that local control makes possible.

Therefore, I do not advocate a single federal test. Instead, I am interested in exploring the alternative of a national examination system. This system would reflect not only the voluntary participation of states, but the powerful combination of their individual expertise. The system would be composed of several examinations anchored by a common set of national standards.

States, which already devote enormous resources to testing, could work in clusters to develop and implement examinations. Each cluster's examination would be designed to assess the range of knowledge and skills encompassed by agreed upon national performance standards.

Clusters might be formed around different interests and states would be free to decide which cluster examination was most appropriate for their use in any given subject. The interests of clusters could evolve over time, and states would be free to change their associations.

Through an anchoring or calibration procedure, a means for equating student performance across clusters could be devised. The key to this process is a commonly agreed upon definition of what students should know and be able to do. These examinations cannot, strictly speaking, be designed until national standards have been developed and adopted. However, we can begin work on the process of designing and evaluating examinations formats. We can develop a format which will work will in the anchoring system and begin to experiment with different methods of calibration.

In suggesting this approach, I do not wish to minimize the degree of technical work that will be called for nor the time required for developing the system. However, at this point, it appears that this type of assessment system is both technically feasible, and essential to meeting our national reform objectives while retaining, and even strengthening, our tradition of local control.

Motivating Changes in the System. The changes that this type of assessment system could help drive include:

- Schools will offer challenging learning opportunities for all students, not just those who are college-bound. For most of the twentieth century, America has had two systems of public education--one designed to train a small group for management and leadership and another to prepare the remainder for the routine work they would do as adults. We now recognize that a work force dominated by individuals who are not challenged to achieve higher than eighth-grade skills is uncompetitive in a greatly changed--and changing world economy.

- Learning environments will be structured to encourage and reward student effort.

- Parents will have the knowledge and motivation to be effective partners with the school because they will understand the skills and abilities their children must have to prosper economically and how their children are performing in relation to this standard.

- The means by which we assess performance will accurately track what we want students to know and be able to do. There is a consensus emerging that the new basic skills include critical thinking, information management, interpersonal and communication skills, and problem solving. Not all classrooms and very few tests presently focus on these skills.

- Assessment of student learning will have consequences. If employers and colleges participate in setting the standards on which the examinations are based, and agree to take performance on the examinations into account when deciding who will be admitted to college and who will get hired at what pay level, then millions of students will, for the first time, see a direct connection between their performance in school and their opportunities in life.

I am a pilot. The system I experienced in obtaining my pilot's license provides a useful model of the kind of system we are aiming toward. When I began my course of study, I clearly know the competencies and understandings that I would have to demonstrate to obtain the license. And I directed my effort at achieving these understandings and competencies. Some came faster and easier than others. The length of my course of study did not depend on an arbitrary number of seat hours, but on mastery of the required skills. I established my command over the required subject matter through a demonstration. The test therefore related directly to the subject matter I was being taught--and to the skills I need to pilot an aircraft safely.

An examination like this one, that can organize and motivate effort, that can accommodate individual differences in learning rates and styles, that provides real consequences for the learner and that establishes standards of performance that all students are expected to meet has much to offer as we reflect on the role assessment can play in education reform.

I welcome a continuing debate on these issues. As chair of the National Education Goals Panel and co-lead governor for education for NGA, I hope and intend to have a close working relationship with this Committee.

Thank you again for the opportunity to participate in this hearing. I would welcome any questions or comments from members of the subcommittee.

Chairman **Kildee**. Thank you very much, Governor. You are a good teacher. I used to use the blackboard myself, and I found it a very effective tool there. I appreciate your presentation.

Senator Bill Brock.

Mr. **Brock**. I am going to begin by departing, Mr. Chairman, from my initial thought to say that I wish every State in the country had as good a governor as Roy Romer. I really appreciate the leadership he is giving to this issue, and I think it is fortunate for the country that he is in the chair of this particular committee. He is articulate, but he is thoughtful, and he is absolutely right. I support, I think, almost every word that he said.

Let me just draw back and try to look at this thing with a bit of perspective. I have been worried about education for a very long period of time, but it got pretty specific when I was the United States Trade Representative and I had to wrestle with people who kept coming in and saying, "Why are we exporting our jobs? Why don't we keep them here? Why don't we have the ability to defend ourselves against people who are outproducing us?" And the first complaint I heard was that all these competitors were coming in with slave wages. Well, that is baloney. We are now not the highest wage country in the world, and we are getting beaten by people who are paying higher wages than we are.

What is the difference? The difference is that we have slipped behind in giving our children and our workers the tools to be as productive as they are capable of being, and that became reinforced when I took the job as Secretary of Labor, and we did this Workforce 2000 study which said, in effect, that unless we change pretty quickly we are not going to be competitive in many areas at all and we will have to compete on the basis of low wages.

Then I went to cochair with Ray Marshall this Commission on the Skills of the American Work Force. We published last year a study called "America's Choice: High Skills or Low Wages," and we came to the conclusion, based upon fact, not opinion, that this country, without understanding it or really consciously deciding to do so, has been trying to compete with the rest of the world on the basis of lower wages rather than higher skills for about 20 years.

Real wages in the United States for 70 percent of the American people are lower today by 12 percent than they were in 1970. Now, doggone it, that has to tell us something. Why is it that, after something like eight years since the "Nation at Risk" was published by the President's Commission back in 1983, we have talked almost ad nauseam about education reform but nothing radical has happened? We have been trying to make changes at the margin.

Why is it that most people know the urgency of education reform but don't think it applies to their school? Why is it always somebody else's problem?

One of the reasons is that we really don't have very good information on what we are doing. One of the reasons is that we get caught up in the shibboleth of politics when we say, "Oh, my gosh, we can't have a national test; that violates local control."

Let me describe what a national test is. An SAT test is a national test. Not everybody is required to take it. College Boards are a national test. Companies give tests; GE gives tests all over the United States; IBM does; they don't vary; they are national tests in a way. The U.S. Army gives a test to every student; it doesn't matter where they come from, what their race or creed is; they all take the same test. Is there something unconstitutional about it? No. What it does is tell the Army who is qualified, or, more importantly, it tells them how to train them to be qualified.

We just went through a war. The Patriot missiles may have been the most important instrument in that war. Well, you can't fire a Patriot missile until you have had, first, a high school education, and then, second, full basic training, and then, third, a full additional year of training in that specific skill.

Well, it seems to me that we need to quit arguing about who has got control of this thing and argue about what we do about a problem. The problem is that our children are not getting an adequate education, and our parents are not being given information on which to base their involvement in the educational system, and if we are going to change that we can't argue about national assessment. As Governor Romer says, we can develop a system of national standards, States can have different methods of testing if they want to, so long as there is a way of relating that test score of their test to the national standard, because we do need to know, and parents need to know, and kids need to know, whether or not they are able to compete globally.

One of the reasons there needs to be a national standard is because we have to benchmark ourselves against the world, not against Texas and Arkansas and Tennessee. We are not competing there, we are competing in the world, and if you look at the purpose of the classroom test, Governor Romer was absolutely right, the primary purpose is to diagnose the problem and let the individual have some understanding of what they have to do to solve it. The purpose is to provide information to the student, the parent, the teacher, so that they can act to solve the problem.

If our medical system were acting like public education is today, you would go to the doctor, and he would treat you and then wait for six months to diagnose you, and then, when you didn't respond, he would say you were the problem.

Well, that is pretty dumb, but that is the way we do things in education, and it seems to me that the first step is to give a different tone to this idea of assessment. I like that word better than "test," by the way. We are the most overtested and undereducated people in the industrial world.

But if we will provide that information that Governor Romer sketched up there, whether it is in math, or science, or in functional skills, like reading a bus schedule, if we provide that information to an individual and let them know where they are in relationship to people around the world that we have to compete with, they will respond.

Those tests allow the student to respond, and they allow us to diagnose the student's problem. Not every student is going to be a ballerina any more than they are going to be an engineer, but each one has to know where they are on a scale of performance in a range of subjects so that they have a sense of what they are trying to achieve. We need those comparisons for the teachers; we need it for schools; we need it between schools and districts and States.

The NAEP does monitor national progress; it does not in any way measure individual students. We need to advance the curriculum, because it is not accurate to say that 50 percent of our 18-year-olds can't read at the ninth grade level and 50 percent of our 18-year-olds can't perform at eighth grade math level.

What we are trying to do in the SCANS Commission is to respond to these problems by suggesting that there is a different way, we can begin to teach within a functional description, we can teach in a real world circumstance, and we can begin to evaluate students like Boy Scouts measure people who are trying to become Eagle Scouts, with a series of competency evaluations over a period of years. If you do that, then you are really beginning to give them the tools with which to achieve the level of progress that they have got to achieve.

What we are looking at is the possibility of trying to have assessment certificates, if you will, or merit badges, if you want to use that term, that would deal with a whole range of different functional skills and use those in the sense that they have to relate to the need or their ability to work in business but also to compete in the world. So they have to be internationally benchmarked.

Let me just conclude with one additional point from this Commission on Skills that Ray Marshall and Ira Magaziner and I worked on with a bunch of other very fine people. The essence of the conclusion we reached was this. Our system has to prepare young people by the age of 16—because that is when they can legally drop out of school in most States—to a globally based standard. We said if we can get to that point, then they ought to be given four more years in what I call a GI bill for all students that would be paid for, in my judgment, by a 1 percent surcharge on the income tax. There are a lot of ways to do that, but that is not the issue. The point is that we ought to give them additional education beyond that in a whole range of fields, but we have given them enough schooling by that time to have a choice that they don't have today.

Secondly, if they don't get to that point, it is not their fault, it is ours, so we ought to have a system of alternative learning centers, because not all students are going to make it in public schools as they are presently described that are available for any student that wants to drop out or choose an alternative method, so that we can get them to that globally derived standard so they can be productive as human beings, and if we do that, then we are beginning to respond to their need and to this country's desperate need to provide our workers with the skills and the tools to be productive internationally in the sense that we can compete with anybody. If we do that, we can eat anybody's lunch and we can start competing on the basis of higher wages and higher skills rather than low wages, which is a suicidal way to go.

Thank you, Mr. Chairman.

[The prepared statement of Senator William E. Brock follows:]

Prepared testimony of

The Honorable William E. Brock
Chairman
Secretary's Commission on Achieving Necessary Skills

Before the

Subcommittee on Elementary, Secondary, and Vocational Education
U.S. House of Representatives
March 14, 1991

Mr. Chairman and members of the Subcommittee:

I appreciate the invitation to participate in this hearing and discuss a very important issue. I have long held a deep and abiding concern about the need to improve the quality of American schools in order to prepare our young people for the realities of work. My experiences as Secretary of Labor and U.S. Trade Representative impressed upon me that our nation's economic future and international competitiveness depends on improving the country's education system.

The landmark report Workforce 2000, which I commissioned as Secretary of Labor nearly four years ago, documented that the skills gap is constraining economic growth in the United States. For nearly two decades, the skills of our workforce have been inadequate and have contributed to a decline in real wages for most Americans and the widening gap between those who go to college and those who do not. Moreover, it has diminished the United States' ability to compete in the global marketplace.

This morning I want to relate the work of the Secretary's Commission on Achieving Necessary Skills, which I chair, to the debate on national assessment. I want to discuss how we are approaching the workforce skills crisis and how we think the issue of national assessment fits into the picture. Let me begin by sharing a few general, personal thoughts about national assessment.

WHAT IS NATIONAL ASSESSMENT?

First, I think all of us who talk about national assessment should clarify what we mean by "national." Alarm bells tend to go off whenever the word "national" is associated with issues surrounding elementary and secondary education. First, we should distinguish between federal and national exams. Examples of national examination systems abound.

Is the College Board, and their SAT and Advanced Placement exams, national? These tests don't vary by state, but the tests are not taken by all students. Are company employment tests national? Again, the tests given by corporations like IBM and GE vary by company, but they don't vary geographically.

Our troops are now coming home after doing a remarkable job in the Persian Gulf. They offer a good example of what I'm talking about. While their high school educational experiences were all over the board, they all took the same tests to get into the armed forces. The Army does not use different enlistment tests in Flint, Michigan and Chattanooga, Tennessee. The Army is able to know how the young man or woman from Flint compares to the new recruit from Chattanooga.

PURPOSES OF ASSESSMENT

Second, we need to understand the purposes and value of assessment. As I see it, the assessment process in education can serve five main purposes:

First, classroom tests can help guide instruction for individual students. Used early and often, they can help students -- and their teachers -- identify problems and address them. If the way most schools use assessment today were transferred to the medical world, doctors would prescribe treatment first, wait five months to diagnose the problem, and then blame the patient because he or she is still sick. In the American classroom, we need to diagnose the problem first and then prescribe a cure and monitor the progress.

Second, assessments and evaluations can be used to identify superior or inferior performance by teachers, schools, school districts, and states. Statistically valid samples are adequate for this evaluation purpose.

Third, systems that have certain consistency to them can help us monitor national progress. The National Assessment of Educational Progress (NAEP) is an example.

Fourth, assessments can be used to influence curriculum, both how it is designed and how it is used. Instead of a test having no relationship with curriculum, it can be used to shape what is being taught. Today, to a degree, Advanced Placement exams do this for the minority of students who participate.

And fifth, assessment systems can certify the competency level of a student so an employer or college can make a reasonable decision about hiring or admission and placement. In a world where students often go to college or work far from home, these exams must be nationally comparable to be useful. The SAT and

Advanced Placement exams serve that purpose for colleges, but there is no similar model for employers. For those who are skeptical about assessment for certification purposes, a voluntary system that is of value to the students should not be a problem.

SCANS AND ASSESSMENT

That brings me to SCANS. Although we have much more work to do, our recommendations on assessment will probably be designed to serve the last two purposes -- influencing curriculum and certifying mastery of skills.

SCANS is based on the premise that today there is an insufficient connection between education and work. Students have little understanding about how school relates to their later work on the job. And they are not being taught the kinds of skills they need for productive employment in today's workplaces. At the same time, businesses are not demanding skilled workers and they are not telling educators, parents, and students what kinds of skills they need. As a result, the value of the high school diploma has diminished. And American youngsters study less than their counterparts in other countries.

SCANS is trying to forge a strong link between school and work. To do this, we believe a fundamental overhaul of the American education system is needed. The change we envision would help provide all students with the competencies they will need for success in the workplaces of today and tomorrow.

Last summer the Commission on the Skills of the American Workforce, which I co-chaired, stressed the need to restructure school and work. This, we felt, is especially important for the so-called "front-line" workers, the majority of people who will not get a college degree. Building on the work done by that Commission, SCANS is now working to recommend specific elements of that restructuring.

To begin, SCANS is identifying the specific skills people need for productive and meaningful employment. These will include the basics of reading, writing, and math, as well as more complex skills, such as the ability to communicate, work in groups, solve problems, and use technology.

But how do we get teachers to teach and students to learn these skills? SCANS is considering the idea of using a series of certificates that will accurately reflect mastery of specific skills.

The certificates would serve the two purposes lacking in our current assessment system. They would provide a clear target for instruction and they would let employers know about the achievement of prospective employees.

For this kind of assessment system to work, it must be part of a comprehensive approach that includes curriculum, instruction, and student motivation. Schools could design curriculum to help students meet certificate requirements. And the certificates would provide a real incentive for students to work hard in school.

It is important for me to stress that SCANS is not talking about a series of multiple-choice tests. More sophisticated forms of assessment are needed to evaluate the acquisition of the complex skills needed in today's workplace. After all, how can a multiple-choice test tell how well an individual works in a group?

Lauren B. Resnick, the Director of the Learning Research and Development Center at the University of Pittsburgh, chairs SCANS's Assessment Committee. Dr. Resnick describes three broad types of assessment that could serve a diverse population and recognize many kinds of skills. They are performance examinations measuring specific tasks, reviews of portfolios of work done over time, and evaluating projects done individually or in groups.

It is also important to stress that the certification system we envision would not be pass or fail. Rather it would measure various levels of achievement, in a similar way that the College Board's Advanced Placement exams have various levels. Our goal is not to identify failures, but to identify problems and help students succeed.

SCANS also is not talking about a single national curriculum. While the certificates would be consistent nationwide, various kinds of curriculum could be used to help students attain them. But all students would be moving toward uniform, high-skill targets.

EMPLOYER BOARD

For such a certificate and assessment system to work, we will need some kind of a governance mechanism. It won't work in a vacuum. And we must ensure that the skills we are demanding are up-to-date, reflecting today's needs.

To ensure that education is effectively linked to work, employers must have a strong voice in determining what skills and knowledge are needed. At tomorrow's SCANS meeting, we will be discussing whether an "employer board" would be the best way to administer the program and issue the certificates. The board would be composed of representatives from business, labor, education, and citizens' advocacy groups. It would be charged with setting objectives and criteria for certificates and establishing a system to assess them.

TAKING ACTION

The ideas I have mentioned describe the direction in which SCANS is going today. But it is important for me to stress that our work is a work in progress. The 31 representatives of business, education, labor, and government who serve on SCANS are reviewing all of these concepts. We intend to release an initial report in late spring and further recommendations later on.

But the bottom line is that SCANS's goal is to produce recommendations that will lead to a fundamental change in the way American schools prepare students for the world of work. The time for rhetoric is over; the time for concrete action is here.

My thinking on these issues has changed dramatically since I served in this House more than two decades ago. But the world has also changed dramatically. The demands of the modern workplace are more complex and we face stiffer competition than ever from abroad. We can no longer be satisfied with tinkering around the edges of an education system that is failing to provide the kind of skilled people our economy needs to remain competitive in the years ahead.

The SCANS mission is ambitious, but the rewards for our success will be substantial. It will lead to a more competitive and vibrant economy, better employment opportunities, and a higher standard of living for our children.

I am pleased that you are taking the time to seriously debate the issue of national assessment. After years of considering how to improve our education system so it better serves individuals and the nation's economic future, I am convinced that some kind of national assessment system is needed to meet that goal.

# # #

Chairman **Kildee**. Thank you very much, Senator Brock.

Dr. Anrig.

Dr. **Anrig**. Thank you, Mr. Chairman.

First, I want to commend you and the members of the committee for having such a hearing. I think what you are doing is shedding light on a complex issue where people are prepared to run with a simplistic answer, and I really admire the opening statements of you and the members of the committee on this issue.

First, Mr. Chairman, I want to be sure to point out to you that I am a graduate of Western Michigan University and I practice taught in Paw Paw, Michigan, and my three roommates in college were from Flint, Michigan. That just goes to show how long I have been preparing for this hearing.

Chairman **Kildee**. You have great credentials.

Dr. **Anrig**. I come here today to speak in opposition to a national test. Such a test, in my judgment, is a well intentioned but ill advised solution. This may seem surprising to you, to hear this from the president of Educational Testing Service, but I take seriously our charter as a nonprofit organization to promote public understanding of tests and of proper test use. I think a national test is not a good way to use tests and is not a proper way to use tests.

I oppose national test proposals for three reasons. First, embracing a national test is like choosing a referee before you know what game you are playing. We first need to define what the educational goals mean in terms that are useful for instruction. I think the national education goals were a very positive first step of the governors and the President. But one of those goals says that by the year 2000 the United States should be first in student achievement in mathematics and science in the world.

Now what does this mean to a fourth grader in Michigan, Pennsylvania, or in the other States represented on this panel? And what does it mean to their teachers? What do we expect them to know and be able to do? We need to answer that question first, before we turn to the question of how to measure their progress. This essential first stage shouldn't be shortcut by letting a test provide the definition. By its very nature, any test is simply a sample of knowledge and skills. If you use the test to define the curriculum, you will end up with a sample curriculum. That doesn't make sense and would not be good for education.

A second reason why I oppose a national test is that, again, as throughout the period of education reform, we are bypassing teachers. Their feeling, in all of the surveys that I have seen of teacher attitudes towards education reform, by the Gallup Organization, the teacher organizations and others, is that reform is something that is being done to them and their students, not for them.

Now, if you take a look at what you are trying to accomplish, what we are all trying to accomplish, in reform, we are trying to achieve improved student learning. As Governor Romer so well pointed out here, you want to be sure that line is going up and that the pilot is not just average—or brain surgeon, just average—but really good in what they are doing. That is what the public is expecting of us. The only people that can deliver that, with all respect to this panel and to us and others in this room—the only people that can deliver that are the teachers in the classroom once that door closes, and we are excluding those teachers from the process of reform.

Some of the business leaders whom I admire very much and have been very active in the reform movement have learned the hard way that competitive productivity means that you involve your work force in deciding what it is they do rather than dictating to it. We can't dictate to teachers and expect them to close the door and develop excellence in their students unless we give them a chance to be part of that action. I think that we are making a bad mistake at this point in not involving teachers more in a central role in reform generally and particularly in the issue of setting standards, determining learning outcomes, and deciding best how to assess and measure them.

Finally, a national test simply won't work, and we have plenty of experience to show this. States and school districts have tried the testing route to educational reform, and it hasn't worked, and a recent report funded by the Ford Foundation said that the amount of effort put into testing right now, today, is 20 million school days per year and over $100 million in direct expenses of tax money right now. Thirty-eight States had statewide testing programs by the end of the 1970's—1970's. Forty-seven States now have statewide testing programs.

What are the results? The National Assessment of Educational Progress, which ETS is privileged to administer on behalf of the National Council for Educational Statistics—NAEP—indicates that since the 1970's, during all this period of all this testing at the local and State level, we have experienced some modest gains in basic skills, some modest gains, and they are quite modest, and only mediocre performance at the middle and higher levels of achievement in reading, mathematics, science, writing, and civics. I see no reason to believe that a national test can do what all this effort and investment by States and localities has failed to do with all of the testing that has gone on up to now. The bottom line for the public is better learning, and we can't test our way to educational excellence.

Now what can work? I believe Governor Romer in the National Goals Panel is on the right track, and I think the SCANS Commission is heading in the right track. The national test would be the wrong track. We have a need for what the national panel refers to as a nationwide assessment system. Now what does this require?

This requires, first of all, that we establish standards for all subjects at important transition-based points, based on a broad-based consensus that includes skills as well as knowledge and address, as the SCANS Commission is suggesting, employment relevant as well as academic outcomes for all students.

Second, we should define what is to be learned before deciding how to assess this learning, and we should focus primarily on outcomes so that teachers have the opportunity for flexibility and creativity in terms of how to best achieve these outcomes.

If you wonder whether that can be done, by the way, the National Council of Teachers of Mathematics and the Mathematics Association of the United States have issued standards for mathematics instruction in the schools. This one publication, not thick, covers all of math from kindergarten through twelfth grade, highly respected standards, without telling people how to do it. It can be done, and it has been done by that organization, and I hope others will do it as well.

We should assure that students are taught what the standards call for. One of my greatest worries about a national test is, we will go off happily and issue a national test and we won't be sure that the youngsters are being taught what is called for in that national test. Now what is that in terms of fairness? That is not fair. That is not good test use.

We should be sure that the students know what the standards are and are taught that, the teachers know what outcomes are expected of them, and that schools know what goals they should strive for, and, as I have said before, I believe teachers should be significantly involved, along with others, in determining each of these.

We should link local creativity and assessment with State and national indicators of performance—this is another part of what I appreciate of the governor's proposal here of clusters—and promote the use of a variety of assessment techniques that include student produced results based on group as well as individual effort. Most tests are designed for individual performance. In most work, you don't work individually, you work with others, and we ought to find ways to begin to build that work habit into what we assess in education.

We should vigorously guard against creating a new system to sort and label children, especially those born into poverty who start childhood with educational disadvantages not of their own making, and, Congressman Hayes, I notice that you spoke to that in your opening comments.

One guard against that is that we should gather and report individually identifiable student scores only when that information will be used to provide improved educational opportunity and/or additional instruction to these students whose scores are to be reported. All other assessment should be conducted on a sampling basis to guard against test misuse. And, finally, we should use assessment positively as a basis for substantial incentives to schools and school districts and, for those in need, substantial help to them.

Now this doesn't describe a national test. What this describes is a national assessment system, and that, Mr. Chairman, is what I came here to speak to today. We need to build on the American traditions, not on those of foreign countries. We could just turn to Europe and Japan and say, well, let's do it their way, but we are not and never have been Europe and Japan. Well, at one time we were partly Europe. But we don't have a centralized system, and, thank God, we don't have a federalized system, and we shouldn't try to make one. We have a diversity of student population, a diversity of strengths, of States, and a commitment to equal opportunity for all kids. That is something we should preserve, and we shouldn't move to a test in a way that will defeat that.

I have identified some existing models in my written testimony that could provide building blocks for a nationwide assessment system. It can be done in parts, it already is being done, and I hope we show good sense in doing that.

Thank you, Mr. Chairman.

[The prepared statement of Dr. Gregory R. Anrig follows:]

Testimony for the Subcommittee on
Elementary, Secondary and Vocational Education
Committee on Education and Labor
U. S. House of Representatives
March 14, 1991

Gregory R. Anrig
President
Educational Testing Service
Princeton, N. J.

National Test: "Nay" - Nationwide Assessment System: "Yea"

There is the beginning of a positive shift away from what has been a bandwagon in favor of a national test to improve American education. I support such a shift. Today I want to outline why a national test will not achieve its intended purposes, and then describe actions that I believe would improve learning for all children.

It may seem strange that the head of the country's largest test development organization is opposed to a national test. One of the central missions for which Educational Testing Service (ETS) is chartered as a nonprofit educational organization, however, is to promote public understanding of tests and of proper test use. I take that responsibility seriously.

Why Not a National Test?

Embracing the idea of a national test to achieve educational excellence is like picking a referee before you know what game you are playing. We first need to define what we expect students to know and be able to do. The National Education Goals announced by President Bush and the nation's governors in 1990 was a good first step. But what does a goal for U. S. students to be first in the world in mathematics and science achievement by the year 2000 mean to a fourth grader in Michigan or Pennsylvania and to his or her teacher? What do we expect them to know and be able to do in order to achieve this goal? These goals need to be translated into terms that have meaning for instruction, by subject and level.

Such terms should not be defined by a test. Any test, by its very nature measures only a sample of knowledge and skills. If tests are used to determine the curriculum, you will end up with a sample curriculum. Effective instruction focuses on what students should know and be able to do, not on what is on a test. Effective achievement tests are valid and reliable measures that focus on what is taught. First decide on your expectations and standards for student achievement. Then decide how best to measure this achievement.

Simply imposing a national test on schools to jump start reform will aggravate even further the general feeling of teachers that education reform is something being done to them and their students rather than for them. Private enterprise has learned the hard way that competitive productivity requires involving the workforce rather than dictating to it. Better student achievement must be developed in the classroom and only the teacher can do it. Teachers should have a central role in reform if we are to benefit from their dedication and commitment. To date, this hasn't happened; a national test would represent another step in the wrong direction.

No Shortcuts to Excellence

Obviously, I am not against tests. I am against using tests in ways that won't work and are educationally unsound. The country has had considerable experience with imposed testing, yet our educational problems persist. A recent report funded by the Ford Foundation estimated that 20 million school days and 100 million tax dollars (for direct expenses) are invested annually in public school testing.

States and school districts already have tried the testing route to reform. By the end of the 1970s, 38 states had statewide testing programs; by the end of the 1980s, this number had increased to 47. While there have been some modest gains in basic skills achievement since the 1970s, the National Assessment of Educational Progress reports only mediocre performance at the middle and higher levels of achievement in reading, mathematics, science, writing, and civics. Other national indicators of student achievement indicate the same plateauing, especially at upper levels of proficiency. Despite earnest efforts and enormous investment of tax dollars in state and local testing programs, student performance still falls far short of what is needed.

Eight years after the release of *A Nation at Risk,* few 12th grade students can demonstrate high levels of performance in reading, mathematics, and science. Few reach a broad understanding of American institutions of government and are able to interpret U. S. historical information and ideas. Few display an ability to produce elaborated responses to various writing tasks.

The widespread public concern that led to this period of education reform was fueled by the belief that American children were not learning well enough for the world that awaits them as adults. The public was right in the 1970s; it was right in the 1980s; and it still is today.

The record is clear that we cannot test our way to educational excellence. Tests alone describe the problem; they don't solve it. A national test is no shortcut to the real reform we all are seeking.

Criteria for an Effective Nationwide Assessment System

If a national test won't work, what will? Let me offer nine criteria for an approach that I believe would work:

- Expand the concept of educational accountability to include states and communities, as well as school districts and schools. Accountability cannot just rest with educational institutions; it must be shared by those responsible for their governance and funding.

- Establish standards for all subjects at important educational transition points, based on a broad-based consensus, that include skills as well as knowledge, and address employment-relevant as well as academic outcomes for all students. We don't just need higher standards, we need new standards equal to the demands of the complex, competitive, technological world environment that faces today's students.

- Define what is to be learned before deciding how to assess this learning. Focus primarily on outcomes so that teachers have opportunity for flexibility and creativity in how best to achieve these outcomes. Provide for levels of performance within standards so that progress can be identified and encouraged, not just success or failure in reaching a single, fixed point.

- Assure that students are taught what the standards call for, that teachers know what learning outcomes are expected of them by subject and grade level, and that schools know what goals they should strive to achieve. Involve teachers significantly, along with others, in determining these standards, outcomes and goals, and how best to assess them.

- Link local creativity in assessment with state and national indicators of performance. Most assessment should continue to be devised locally for instructional purposes. Each level of government should stimulate assessment ideas and training in assessment for the educational system. Promote the use of a variety of assessment techniques that include student-produced results based on group as well as individual effort. Since standards can never cover everything we wish children to learn, encourage diversity of learning, cultural understanding and enrichment within the context of common standards.

- Vigorously guard against creating a new system to sort and label children, especially those born into poverty (24 percent of those under age one in 1989, according to the Bureau of the Census) who start childhood with educational disadvantages not of their own making.

- Gather and report individually-identifiable student scores only when that information will be used to provide improved educational opportunity and/or additional instruction to these students. All other assessment should be conducted on a sampling basis to guard against test misuse.

- Use assessment positively as a basis for substantial incentives to schools and school districts for progress towards achieving outcomes and standards, and for substantial help to schools and school districts whose students are not progressing satisfactorily.

- Design the system so it is dynamic and can be improved over time based on experience and advances in knowledge.

These criteria do not describe a "national test". They describe a nationwide assessment system that builds upon the traditions of American education, with its diverse student population, rather than imitating the highly centralized (and federalized) systems of Europe and Japan. An interim report of the National Education Goals Panel, chaired by Colorado Governor Roy Romer, recommends creating such a nationwide assessment system by the end of this decade. The panel is on the right track. A national test is the wrong track for education and for the country.

Promising Models for Success - Existing Elements

Elements of such a nationwide assessment system already exist. They can be learned from and built upon. Let me describe some of them:

National Council of Teachers of Mathematics, American Association for the Advancement of Science, National Research Council's Mathematical Sciences Education Board. These three organizations are setting new standards and learning outcomes that are earning the respect and ownership of those who must carry them out. They also are involving a broadly representative group in identifying promising ways to assess the new standards and outcomes.

State Initiatives for New Ways to Assess Learning. States such as California, Connecticut, and Vermont are pioneering new forms of assessments based on state-developed learning outcomes. In each case, teachers have had a central role in the initiative.

The National Assessment of Educational Progress and Trial State Assessments. The National Assessment is attracting increased attention and credibility as "The Nation's Report Card". It has an invaluable capacity for tracking performance trends over time and future progress towards the new National Education Goals. The National Assessment also is becoming a vehicle for developing measurement innovations that are technically and economically feasible for large scale assessments on a nationwide basis. Approximately 25 percent of the items in the 1990 assessment were performance-based and this proportion will increase significantly in the 1992 assessment. In addition, thirty-seven states are participating voluntarily in the Congressionally-authorized Trial State Assessments this year. Only one of these states has dropped out of the Trial State Assessments for 1992 and four new states have signed up to replace it. In June of this year, the first assessment reports will be issued that will permit valid state-by-state comparisons of student performance in mathematics.

Resnick/Tucker Proposal for a National Examination System. Unlike proposals for a national test, Resnick and Tucker propose a national examination system that is syllabus-based, voluntary, and draws upon a variety of assessment techniques (examinations, portfolios, projects) that will protect flexibility and creativity for teachers at the local level.

The College Board's Advanced Placement Program. Operational nationwide for 35 years, the Advanced Placement Program provides 29 examinations in 15 subject matter fields. More than 330,000 high school students took Advanced Placement examinations in 1989-90, an increase of 175 percent since 1980. These respected examinations are syllabus-driven, are largely performance-based (essays, portfolios, problems), and are nationally judged by carefully trained assessors who are high school teachers and college professors. A total of 9,292 high schools across the country provide Advanced Placement courses and 1800 colleges and universities use Advancement Placement grades to determine course placement and credit. The number of urban high schools and minority students participating successfully in the Advanced Placement Program has increased dramatically in the 1980s.

National Assessment Governing Board Standards. The Governing Board has begun the complex task of developing achievement levels for subjects and grades assessed by NAEP. While I have expressed concerns to the Board about the procedures used for its initial effort, I support the development of standards at several proficiency levels for the National Assessment, and the Governing Board's role in doing so.

The College Board's Educational EQuality Project. The College Board involved literally thousands of educators and others in developing its "Green Book", *Academic Preparation for College*, and the follow-up "Rainbow Books" that describe high standards of expected accomplishment in each academic area for college-bound students. The process followed by the College Board demonstrated one viable model for developing a broad-based consensus on learning outcomes.

Towards A Common Curriculum and Instructionally Useful Assessment - Two Remaining Issues

I advocate achieving a consensus on standards and learning outcomes before deciding on how best to assess student performance.

Wouldn't this result in a "national curriculum" feared by so many educators and others over the years?

I always have felt the issue of a national curriculum was something of a red herring. If one were to take a helicopter journey across the country, touching down every 150 miles, and walk to the nearest school, I believe you would find more curriculum similarity than difference among schools nationwide. This is largely due to textbooks marketed on a national basis by a relatively few commercial publishers. The problem with this is that a common curriculum is being defined de facto by vagaries of the textbook marketplace.

I am coming to the personal opinion that educators need to reassert responsibility, as has the mathematics profession, and organize with others a deliberate effort to determine a common curriculum for public education in the United States. Such a common curriculum should center on what should be learned, giving teachers and schools flexibility to be creative in how to achieve this learning. American Federation of Teachers President Albert Shanker proposed the need to develop such a common curriculum in his February 24, 1991 weekly column in the New York Times. I agree with him.

An area in which my ideas are still forming has to do with what is called instructional assessment. Traditionally, standardized testing in the United States has been expected to serve two purposes -- public accountability and instructional improvement. While these two purposes are not totally incompatible, I am becoming increasingly convinced that the use of one instrument for both purposes may result in neither purpose being served as effectively as it should. The ETS Board of Trustees annually publishes (as part of the ETS Annual Report) an Accountability Report on matters of public concern. Its 1990 Accountability Report is titled "Instructional and Accountability Testing in American Education: Different Purposes, Different Needs". I append this report to my testimony because of its relevance to the issue of a national test.

I thank the Subcommittee on Elementary, Secondary and Vocational Education for the opportunity to present my views on this important subject.

INSTRUCTIONAL AND ACCOUNTABILITY TESTING IN AMERICAN EDUCATION:

DIFFERENT PURPOSES, DIFFERENT NEEDS

TESTING IN AMERICAN EDUCATION

Each year, the ETS Board of Trustees prepares a public accountability report on a matter of public concern. This year, the board chose to report on issues related to instructional and accountability testing.

Educational testing will change more in the next 10 years than it has in the past 50 years. This change will be driven by continuing pressures for better educational results – such as by the new National Education Goals announced in 1990 by President Bush and the nation's governors – and by promising advances in cognitive psychology, technology, and measurement theory. These advances make possible dramatic new forms of educational assessment in the 1990s and beyond.

Such change will significantly affect the two distinct and sometimes conflicting functions of testing in American education – public accountability and instructional improvement.

Accountability Testing

Public officials, employers, and taxpayers demand accountability from school officials for the billions of tax dollars invested annually in schools, and for the educational impact of that investment. As one response to this demand, accountability testing has been designed to measure a sample of what large numbers of students have learned, and what skills they can perform, at various points in their education.

Instructional Testing

Meanwhile, teachers, school administrators, and parents demand assessments that can guide the improvement of teaching and learning. Instructional testing is designed to give teachers and students information that will help them in the classroom so they can meet rising expectations for better academic achievement. Ideally, such assessment should itself be a vehicle for individual learning, rather than an impersonal snapshot of a national or regional sample.

Different Purposes, Different Needs

It has been difficult for these two purposes – public accountability and instructional improvement – to be served effectively by the same kind of test. In the future, however, it is increasingly possible that each kind of measurement can be enriched by what is learned from the other, in terms of both test design and test results.

ETS is committed to being a leader in developing new approaches to instructional and accountability assessment. (See the 1990 ETS Annual Report for a more detailed description of these new assessments.) It is hoped that these new approaches to assessment may even broaden how taxpayers think about accountability and what they value in terms of educational results.

Public Concern for Accountability Testing

Concerned taxpayers justifiably want to know how good a job schools are doing. They want students to learn better, and to learn more, so these students can be equal to the demands facing them in the 21st century and can meet the competition from other nations that seem to have more rigorous schooling. One response to these concerns has been the use of standardized tests to inform the public about how schools are performing. These tests have certain features in common:

- They have been imposed from above by a public authority such as a school district or state legislature.
- Since these tests have been funded primarily from tax revenues, policymakers have relied heavily on multiple-choice, machine-scored tests in order to hold down the costs.
- Because their primary audiences are taxpayers and policymakers, such tests are designed to yield summaries of performance at the school, district, and state levels. These summaries too often have been used simply to create rankings or to support other superficial judgments about educational quality.

Educators' Concern about Accountability Tests

Teachers and other educators contend that accountability tests do not address some of the most important things they are trying to teach and do not help them to improve instruction in the classroom. They also complain that the amount of mandated testing is consuming too much instructional time. To some educators, these snapshot samples of what is being learned distort rather than describe the breadth and depth of what they are trying to teach.

As accountability tests became more widespread in the 1980s, concern also increased that schools were responding to accountability pressures by teaching only to the subject matter

---

In this report, "tests" is used to refer to standardized educational tests as we now know them. "Assessment" is used to describe measurement of broader skills in new ways that are more useful for teaching and learning.

covered in the tests. Thus, the tests ended up defining the curriculum to be taught and became the primary focus of what was done in the classroom. Existing accountability tests are inappropriate models for this.

Balancing the Concerns of Accountability and Instruction

There is a need for new kinds of assessments that represent desired educational goals and contribute in constructive ways to the teaching and learning process. Many see the use of well-designed performance tasks in tests (along with or instead of multiple-choice questions) as bringing us closer to these aims. We share that view.

It is important, though, to recognize the extent to which performance measures differ from existing accountability tests. When used for instructional assessment, they must be closely tied to the specific curriculum being taught in the classroom. They require substantial student and teacher time to accomplish and they require carefully trained assessors to evaluate the results. All of this represents a costly investment. To justify that investment, there must be clear evidence that such assessment significantly informs instruction and enhances academic achievement.

Such an investment is probably too great for general accountability purposes. For example, one state with a strong commitment to educational assessment found that redesigning its state program around performance tasks would increase by tenfold the cost of the existing state assessment program. All states face competing priorities for scarce education dollars and there just aren't enough resources to do all that needs to be done.

ETS and the Two Types of Assessment

ETS is involved in an intensive development effort that may help reduce the gap between instructional assessment and accountability assessment, and may improve both in the process. For accountability purposes, we are developing performance measures that would be relatively inexpensive to administer on a large scale. For instructional assessment, there are research and development projects now under way to construct instruments that will both evaluate and facilitate student learning. These efforts should, in time, yield more economical and practical forms of assessment to promote learning.

National Assessment. A current example of this more balanced approach for accountability testing is the National Assessment of Educational Progress (NAEP), which ETS develops and administers for the U.S. Department of Education. Combining multiple-choice and performance measures, NAEP is demonstrating the feasibility of a large-scale assessment that includes performance tasks. NAEP is economically feasible, however, because it is administered only to a sample of students at three grade levels.

Technological Advances. Technological advances make possible new ways of measuring increasingly complex performance skills. For example, computers can be used to score a multi-step mathematical problem. A computerized analysis of the results from such a task can also provide important information to teachers on where students encounter difficulty in solving problems. Since students can make changes to images on a computer screen, they also can make changes to solve realistic simulations of challenging tasks and problems portrayed on the computer screen. Working for the National Council of Architectural Registration Boards, ETS has developed computer performance tasks that simulate actual problems encountered in designing buildings. In other instances, paper-and-pencil tests are being computerized to offer students the chance to demonstrate their knowledge and skills more efficiently and to provide them with an immediate report of the results.

Instructional Assessment. As part of a major commitment to create a new generation of assessments designed specifically to improve student learning in the classroom, ETS is developing instructional assessments for critical thinking, science, mathematics, writing, and the arts for the middle school level. Several elements characterize these projects:

- a blurring of the line between instruction and assessment,
- careful involvement of teachers in the design of the instructional assessments, and
- extensive field testing in classrooms with diverse groups of students to determine the impact of the instructional assessment on student learning.

These projects may not yield performance tasks that can be easily adapted for large-scale accountability purposes. But they already are providing valuable insights into the strengths and limitations of both instructional and accountability assessments — and what can be learned from each that might benefit the other.

What Needs to Be Done

The same kinds of tests have been used in this country for quite different purposes. Experience with the multiple-use approach over the last two decades clearly demonstrates its limitations, and highlights the need for a closer link between purpose and the type of assessment chosen to serve that purpose.

We also must understand that tests designed for one purpose may have a major — and often unintended — impact in other arenas. The carry-over effect is one of the vexing issues facing accountability testing. Although tests provide useful information about what students have learned in school, they also can narrow, in undesirable ways, the very content of the learning schools seek to achieve (as with teaching to the test).

If we can bring these insights to bear more fully in the public dialogue about educational testing, then we will have taken a major step toward the proper use of assessments for the betterment of all participants in the learning process.

The United States is trailing in the international race for greater competency. If we are to regain a leading position in this competition, public education needs greater public confidence and better educational results. ETS is fully committed to contributing to both by creating new kinds of assessments that inform and help instruct so that high educational goals can become attainable for all children in the United States.

Prior Trustees' Accountability Reports

1989	Toward a Better Understanding of Gender and Testing
1988	Common Sense on Preparing for an Admissions Test
1987	Addressing Issues of Test Bias
1986	Teacher Standards and Teacher Testing: A Fast-Changing Focus of Educational Reform
1985	Development of Educational Tests in the United States: Privately Directed, Publicly Responsible
1984	Excellence, Equity and Testing in a Period of Educational Reform
1983	Promoting Proper Test Use
1982	Shaping a Corporate Commitment: Openness in Testing

For copies of selected ETS publications and additional copies of this 1990 Public Accountability Report, please write to Communication Services, Mail Stop 47-D, Educational Testing Service, Rosedale Road, Princeton, NJ 08541-0001.

Educational Testing Service
Helping America
Understand educational needs
Reach educational standards

Educational Testing Service, ETS, and ® are registered trademarks of Educational Testing Service.
Educational Testing Service is an Affirmative Action/Equal Opportunity Employer.
Copyright © 1990 by Educational Testing Service. All rights reserved.

EDUCATIONAL TESTING SERVICE

1990 BOARD OF TRUSTEES

Harold Howe II, Chair
Senior Lecturer
Graduate School of Education
Harvard University
Cambridge, Massachusetts

Reatha Clark King, Vice Chair
President and Executive Director
General Mills Foundation
Minneapolis, Minnesota

Gregory R. Anrig
President
Educational Testing Service
Princeton, New Jersey

Tomas A. Arciniega
President
California State University, Bakersfield
Bakersfield, California

Terrel H. Bell
Professor of Educational
Administration
University of Utah
Salt Lake City, Utah

Paula P. Brownlee
President
Association of American Colleges
Washington, DC

Yvonne Brathwaite Burke, Esq.
Attorney
Jones, Day, Reavis & Pogue
Los Angeles, California

Alan K. Campbell
Visiting Executive Professor
The Wharton School
University of Pennsylvania
Philadelphia, Pennsylvania

Jack G. Clarke
Senior Vice President and Director
Exxon Corporation
Irving, Texas

Norman C. Francis
President
Xavier University
Houston, Texas

William H. Harris
President
Texas Southern University
Houston, Texas

John L. Loeb, Jr.
Chair
John L. Loeb, Jr. Associates
New York, New York

Robert M. O'Neil
Director
The Thomas Jefferson Center for
the Protection of Free Expression
Charlottesville, Virginia

Manuel T. Pacheco
President
University of Houston-Downtown
Houston, Texas

Sister Joel Read
President
Alverno College
Milwaukee, Wisconsin

Terry T. Saario
President
Northwest Area Foundation
Saint Paul, Minnesota

Gerald N. Tirozzi
Commissioner of Education
State of Connecticut
Hartford, Connecticut

Chairman KILDEE. Thank you, Dr. Anrig.

Dr. Sewell.

Dr. SEWELL. Mr. Chairman and members of the subcommittee, I am the acting dean of the College of Education at Temple University and a fellow of the American Psychological Association. I wish to thank the subcommittee for inviting me to testify today on the important issue of national testing.

As Americans, we are justifiably proud of the favorable conditions under which the majority of our children are reared. We know of the physical, social, and psychological care provided to those born of advantaged status, and we know of the quality of the intellectual stimulation associated with middle class child rearing practices. We know of the emotional investment of parents in the transmission of the culturally important elements of the society. We know, too, of the care with which children are exposed to a wide variety of experiences. Those of us who are educators, through our research and our experience in the schools, also know how all of these factors can positively affect a child's overall functioning, including the ability to take standardized tests.

But there is another social reality into which a large percentage of America's children are born. It is a world in which their lives are battered by poverty, social isolation, and often racial and ethnic inequities. It is an environment where crime, drug addition, pollution, and other factors are daily realities. There are rural and urban communities where children attend school physically and psychologically unprepared to learn. There are social conditions in which the nutritional status and health care needs of the children adversely affect academic achievement. It is a grim world, far removed from the environment of America's privileged middle and upper class youth. It is a world in which taking and doing well on a standardized test is often irrelevant and sometimes impossible. It is my awareness of these conditions and of their impact on our children that makes me irrevocably opposed to a national standardized high school examination.

Although poverty and a socially disadvantaged background are not insurmountable barriers to outstanding performance on achievement tests, children who experience these social ills consistently manifest lower performance on a variety of standardized tests when compared to children of upper socioeconomic status. There are communities in which lack of opportunities and low school and community expectations are often cited as reasons for an approximate school dropout rate of 50 percent.

The Federal Government is fully aware of these situations. The State governments have reported these statistics based on their own testing programs. The local communities can identify each and every school in which substandard academic performance has become a political football. The local and State governments can also provide evidence that the expenditure per pupil in these low performance schools is usually between one-third to one-half of that of the wealthier communities. The issue, therefore, is not a lack of awareness of the difficulties we face as a nation but, rather, the tendency to propose simple solutions for complex problems. A national graduation exam is this type of a simple solution. To quote from the 1990 report of the National Commission on Testing and Public Policy, "We cannot test our way out of educational problems."

But despite the widely recognized problems with standardized tests, there is still a call for a national test. While there are many reasons cited in favor of a national test, I would like to focus on a few of the major ones and give my reactions to them.

Point one: We need to know what works. This argument usually focuses on the issue of accountability, of finding out which of our schools is doing well and which ones are doing poorly. It is often raised in the context of how tax dollars are spent and what communities are getting for these dollars.

While I recognize why this argument is raised, I have several objections. First of all, a national test cannot tell us anything more than we already know. We already know that poor, urban schools have lower mean scores on standardized tests than middle class, suburban schools. Merely proving this one more time does nothing to remedy the problems that have given rise to this discrepancy.

Secondly, America is already spending over $700 million per year on testing. Although there is no national test, the emphasis on testing at the expense of instructional time has already exceeded reasonable limits. According to a 1985 national survey, 47 of the 50 States mandate some form of statewide testing on student attainment. In my opinion, one of the major lessons we have learned from all this testing is that implementing testing programs has no discernible impact on student achievement. There is no reason to believe that a national test will have any different outcome.

Point two: A national test can provide valuable information for employers, colleges, and other consumers of our educational output. This is another version of the "what works" rationale, except that the focus is on providing a way to differentiate between students at the time they graduate from high school.

I have several concerns about this argument. My primary objection is that this argument runs counter to an essential American value, that change and growth are an inherent part of life and that redemption is always possible. We all know people who, as they grow older, are better able to cope with their difficulties. Some are able to leave a dysfunctional or poor family through hard work in diverse jobs. Others find that they begin to learn how to handle family expectations, overcome learning problems, and in time develop a better sense of worth. What will become of these people if they are effectively screened out of greater possibilities at the time they graduate from high school?

In addition, if we accept the need for some type of accountability at the time of graduation from high school, there is no implied relationship between this accountability and testing for instructional purposes. If excellence in academic achievement is the central goal of a national test, how will the scores enhance achievement and instruction? I think there is ample evidence to indicate that standardized tests do not achieve this objective, as the experience in my own State of Pennsylvania indicates.

The State of Pennsylvania mandated a test to be given to all students in the State to assess the performance levels of schools. Monetary help was provided to support remedial efforts where indicated. Schools in the most affluent segments of the State were compared to those in the most economically depressed areas in which low performance was already documented. The promised resource for remedial help was rather quickly discontinued, yet the comparison persisted. If this example can be generalized to the national level, why would we spend an additional $100 million for testing when the political reality of the local community with the constitutional control over the curriculum and school financial support is unlikely to change?

Point number three: Almost all of the industrialized nations of the world have such an examination. This argument compares the United States with countries such as England, Germany, and Japan, and seems to imply that there is some connection between the economic advantages that these countries are experiencing and the existence of a national exam.

While it is true that many countries have a national exam, to compare and contrast diverse cultures is simplistic and misleading. From a purely pragmatic perspective, the situation in the United States and countries like England and Germany and Japan are not the same. These countries test far fewer students than we do, and they test far more selectively. This allows these countries to use exams which demand integration and synthesis. The possibility of reproducing this type of test at a national level in this country seems remote.

More importantly, countries like England and Japan test their citizens to limit access to higher education and to discriminate among those who will have access to better schools. It is unlikely that any of the proponents of a national test would make this one of its explicit goals. However, I believe that the proposed national test, rooted in the assumption of higher national standards, might simply be a thinly veiled effort to use a scientific rationale to advance an elitist social structure in the United States. If not so, why then, given the rich American tradition of local control over the educational enterprise, should a national high school test dictate employment, a high school diploma, or college admission?

Point number four: A national test will have the desired outcome of showing what works without the negative outcome of leading to a mandated national curriculum. It is difficult to take this argument seriously. Proponents of the national test seem to forget that curriculum and achievement tests go hand in hand. In fact, if they don't, what is the test measuring? Moreover, these proponents also argue that the existence of a national test will not lead to teaching for the test. Again, it is difficult to accept this argument. Already, without a national test, in a recent Gallup Poll, 73 percent of the teachers said that they feel pressured to prepare children for testing; 57 percent said that the testing determine what they taught. Let us be clear about this issue: A national test will inevitably lead to a national curriculum. Even more importantly, invaluable classroom time will be spent preparing for the test.

I would like to summarize my presentation by pointing to the fact that testing has always been a mechanism subject to the political realities of the community. There is a seductive argument that attributes to testing the salutary value of improving the opportunities of low income and minority children by reducing the pervasive class and ethnic bias associated with personal, subjective judgments. The proponents of standardized tests may also believe that the social concept of equality of educational opportunities is best served by the objectivity of tests.

My own belief, based on research and years of experience in the schools, is that differential socioeconomic circumstances make standardized tests inherently biased against poor and/or minority children and, consequently, negatively affect educational opportunities and success.

The advocates would lead us to believe that American educational excellence will lead to the promised land and that one of the vehicles to achieve this objective is to develop a national standardized test. Even if this argument is true—and I hope that I have demonstrated why I don't believe that it is—the question could still be asked: For whom is this supposedly excellent system designed? For many it will not be a promised land but a land of shattered dreams.

Thank you, Mr. Chairman.

[The prepared statement of Dr. Trevor Sewell follows:]

TESTIMONY OF

TREVOR E. SEWELL, PH.D.
TEMPLE UNIVERSITY

before the

COMMITTEE ON EDUCATION AND LABOR
U.S. HOUSE OF REPRESENTATIVES

on the subject of

NATIONAL TESTING

MARCH 14, 1991

2261 RAYBURN HOUSE OFFICE BUILDING

Mr. Chairman and Members of the Subcommittee:

I am Trevor E. Sewell. I am Acting Dean of the College of Education at Temple University and a Fellow of the American Psychological Association. I wish to thank the Subcommittee for inviting me to testify today on the issue of national testing.

A National Test: Opportunities for the Few, Shattered Dreams for the Many

Introduction

As Americans, we are justifiably proud of the favorable conditions under which the majority of our children are reared. We know of the physical, social and psychological care provided to those born of advantaged status, and we know of the quality of the intellectual stimulation associated with middle-class child-rearing practices. We know of the emotional investment of parents in the transmission of culturally important elements of the society. We know, too, of the care with which children are exposed to a wide variety of experiences. Those of us who are educators, through our research and our experience in the schools, also know how all of these factors can positively affect a child's overall functioning, including the ability to take standardized tests.

But there is another social reality into which a large percentage of America's children are born. It is a world in which their lives are battered by poverty, social isolation and often racial and ethnic inequities. It is an environment where crime, drug addiction, pollution and other factors are daily realities. There are rural and urban communities where children attend school physically and psychologically unprepared to learn. There are social conditions in which the nutritional status and health care needs of the children adversely affect academic achievement. It is a grim world far removed from the environment of America's privileged middle and upper class youth. It is a world in which taking and doing well on a standardized test is often irrelevant and sometimes impossible. It is my awareness of these conditions and of their impact on our children that makes me irrevocably opposed to a national standardized high school examination.

Although poverty and a socially disadvantaged background are not insurmountable barriers to outstanding performance on achievement tests, children who experience these social ills consistently manifest lower performance on a variety of standardized tests when compared to children of higher socioeconomic status. There are communities in which lack of opportunities and low school and community expectations are often cited as the reasons for an approximate school drop-out rate of

50 percent. The Federal government is fully aware of these situations. The state governments have reported these statistics based on their own testing programs. The local communities can identify each and every school in which substandard academic performance has become a political football. The local and State governments can also provide evidence that the expenditure per pupil in these low performance schools is usually between 1/3 to 1/2 of that of the wealthier communities. The issue, therefore, is not a lack of awareness of the difficulties we face as a nation, but rather the tendency to propose simple solutions for complex problems. A national graduation examination is this type of simple solution. To quote from the 1990 report of the National Commission on Testing and Public Policy: "we cannot test our way out of educational problems".

Arguments Against a National Test

But despite the widely recognized problems with standardized tests, there is still a call for a national test. While there are many reasons cited in favor of the national test, I would like to focus on a few major ones and give my reactions to them.

(1) **We need to know what works.** This argument usually focuses on the issue of accountability; of finding out which of our schools is doing well, and which ones are doing poorly. It is often raised in the context of how tax dollars are spent and what communities are getting for these tax dollars.

While I recognize why this argument is raised, I have several objections. First of all, a national test cannot tell us anything more than we already know. We already know that poor, urban schools have lower mean scores on standardized tests than middle-class, suburban schools. Merely proving this one more time does nothing to remediate the problems that have given rise to this discrepancy. Secondly, America is already spending over 700 million dollars per year on testing. Although there is no national test, the emphasis on testing at the expense of instructional time has clearly exceeded reasonable limits. According to a 1985 national survey, 47 of the 50 states mandate some form of state-wide testing of student attainment. In my opinion, one of the major lessons we have learned from all of this testing is that implementing testing programs has no discernible impact on student achievement. There is no reason to believe that a national test will have any different outcome.

(2) **A National test can provide valuable information for employers, colleges and other consumers of our educational output.** This is another version of the "what works" rationale, except that the focus is on providing a way to differentiate between students at the time they graduate from high school.

I have several concerns about this argument. My primary objection is that this argument runs counter to an essential American value -- that change and growth are an inherent part of life, and that redemption is always possible. We all know people who, as they get older, are better able to cope with their difficulties. Some are able to leave a dysfunctional or poor family through hard work in diverse jobs. Others find that they begin to learn how to handle family expectations, overcome learning problems and in time develop a better sense of self-worth. What will become of these people if they are effectively screened out of greater possibilities at the time they graduate from high school?

In the past, the special feature of our American educational system has been its emphasis on access to lifelong learning. For example, an older person can return to high school even after dropping out; a community college can serve as a gateway to opportunity; state colleges and universities have provided excellent higher educations at bargain rates; doors have opened for "late bloomers" when they were ready to learn; immigrant children have been able to obtain a college degree. In short, America has continued to be the land of educational opportunity. Will a standardized examination continue this tradition, or will it screen out from further education people from diverse cultures, from different races, of different genders, and of varied ages?

In addition, even if we accept the need for some type of accountability at the time of graduation from high school, there is no implied relationship between this accountability and testing for instructional purposes. If excellence in academic achievement is the central goal of a national test, how will the scores enhance instruction and achievement? I think there is ample evidence to indicate that standardized tests do not achieve this objective, as the experience in my own state of Pennsylvania indicates.

The State of Pennsylvania mandated a test to be given to all students in the State to assess the performance levels of its schools. Monetary help was provided to support remedial efforts where indicated. Schools in the most affluent segments of the state were compared to those in the most economically depressed areas in which low performance was already documented. The promised resource for remedial help was rather quickly discontinued, yet the comparison persisted. If this example can be generalized to the national level, why would we spend an additional $100 million for testing when the political reality of the local community with the constitutional control over the curriculum and school financial support is unlikely to change?

(3) **Almost all of the industrialized nations in the world have such an examination.** This argument compares the United States to countries such as England, Germany and Japan, and seems to imply

that there is some connection between the economic advantages that these countries are experiencing and the existence of a national exam.

While it is true that many countries have a national examination, to compare and contrast diverse cultures is simplistic and misleading. From a purely pragmatic perspective, the situation in the U.S. and countries like England, Germany and Japan are not the same. These countries test far fewer students than we do, and they test far more selectively. This allows these countries to use essay examinations which demand integration and synthesis. The possibility of reproducing this type of test at a national level in this country seems remote.

More importantly, countries like England and Japan test their citizens to limit access to higher education and to discriminate among those who will have access to better schools. It is unlikely that any of the proponents of a national test would make this one of its explicit goals. However, I believe that the proposed national test, rooted in the assumption of higher national standards, might simply be a thinly veiled effort to use a scientific rationale to advance an elitist social structure in the United States. If not so, why then, given the rich American tradition of local control over the educational enterprise, should a national high school test dictate employment, a high school diploma or college admission?

(4) A national test will have the desired outcome of showing "what works" without the negative outcome of leading to a mandated national curriculum. It is difficult to take this argument seriously. Proponents of the national test seem to forget that curriculum and achievement tests go hand in hand. In fact, if they don't, what is the test measuring? Moreover, these proponents also argue that the existence of a national test will not lead to teaching for the test. Again, it is difficult to accept this argument. Already, without a national test, in a recent gallop poll, 73% of the teachers said that they feel pressured to prepare children for testing; 57% said that tests determine what they taught. Let us be clear about this issue: a national test will inevitably lead to a national curriculum. Even more importantly, invaluable classroom time will be spent in preparing students for the test.

Even if we accept the possibility that a national high school test can be developed to meet acceptable technical properties of validity and reliability, what will be the social consequences? This question is of particular importance in light of the absence of a mandatory national curriculum and textbooks, the discrepancy between local communities' per pupil expenditures, and the adverse impact of social-class and racial inequities deeply entrenched in American education.

I would like to summarize my presentation by pointing to the fact that testing has always been a mechanism subjected to the political realities of the community. There is a seductive argument that attributes to testing the salutary value of improving the opportunities of low income and minority children by reducing the pervasive class and ethnic biases associated with personal subjective judgements. The proponents of standardized tests may also believe that the social concept of equality of educational opportunities is best served by the objectivity of tests. My own belief, based on research and years of experience in the schools, is that differential socio-economic circumstances make standardized tests inherently biased against poor and/or minority children, and consequently negatively affect educational opportunities and success.

The advocates would lead us to believe that American educational excellence will lead to the promised land, and that one of the vehicles to achieve this objective is to develop a national standardized test. Even if this argument is true, and I hope that I have demonstrated why I don't believe that it is, the question could still be asked: for whom is this supposedly excellent system designed? For many, it will not be a promised land, but a land of shattered dreams.

Chairman **Kildee**. Thank you very much, Dr. Sewell.

The subcommittee has heard some concurring and some contrasting views on the question of assessment and testing. I would like to ask a general question and have any or all of you respond to that.

Testing or assessment, for the most part, takes place at the end of an experience or the end of a process. What do we discover in testing that can help us to improve that experience or process? As I mentioned yesterday, testing is generally an omega point. What do we do with alpha, beta, gamma, delta? How do we make testing helpful in the improvement of education?

Governor.

Mr. **Romer**. I think that a test ought to do four things. First, it ought to organize one's work; it ought to motivate them; secondly, it ought to accommodate individual differences in learning; third, there ought to be real consequences for the learner out of a test; and, fourth, you have got to have clear standards that you are reaching for. Let me go back to the pilot test again, because it is a very good illustration.

When you are teaching somebody, you teach to the test, and if you have got the right test, it is a good thing. If you have got the right pilot test, that test, as you go through that course of training, can tell you, you don't need to spend your time on this, you have it, but you do need to spend your time on that.

Secondly, there are real consequences to that test which reward someone for having done it right, and this is one of the things that I just think, as I listen to all of us this morning, even though I was not advocating a national test but a national system, I would not like us to conclude that the system as we have it is okay, because we have the level of educational performance in this Nation which is quite inadequate because people think it is adequate; that is the problem. We, frankly, out there think, "Hey, as a nation, we are in pretty bad shape, but our school is doing a very good job," and this wrong, and we need to have assessment systems that make us, as parents and as students, know that we are ill educationally, as Senator Brock has said, compared to what our vocational requirements are going to be when we are adult.

Chairman **Kildee**. Senator.

Mr. **Brock**. I will try to respond in a slightly different direction—I agree with that—but what you can do with tests. I don't think I have heard anybody suggest a national graduation test; that isn't even an issue before us, as far as I know. What I thought we were talking about was a series of tests and evaluations and assessments that allow a student to know where he is in the spectrum of progress and allow us to know what we need to do get them up to the full scale of productivity that they are capable of.

So what a test can do for you—a series of tests or a series of assessments, whatever word you want to use—is to say that by the fourth grade you have gotten to this point and you are not going to make it to the level that you want to achieve by the time you are 16 or 17 unless you change your level of performance in this particular area.

So it allows the student to be given the knowledge of what they have got to do to begin to move in the right direction. It allows the

188

parent to know and allows the teacher to know where to focus their efforts.

Canada has a wonderful program that was developed by private enterprise, funded by government, called Yes, Canada. You can sit down at a computer, and in 30 minutes that computer will analyze you. It doesn't matter whether you are three or 80, it will start talking to you in your language, and, from then on, if you say, "I want to study, off the mainframe data base, Spanish," or English, or math, today, the computer works with you for that hour of instruction at that level, and the teacher is sitting in the room with a master board, and the teacher will all of a sudden see a flashing light, and the computer will say, "Teacher, Bill is having trouble with his algebra; go over and work with him as an individual." Then you are beginning to use the teacher's real talents on an individual basis. We don't do that today. We lecture, we hector, and we are terrified of evaluation and assessment.

You can use those assessments, I think, to give the students the tools they have got to have to make that progress and to give the teacher what they simply have to have. We are not giving our teachers any clear message today. They know what it takes to send a child to college. They have no idea whatsoever what it takes to prepare a child for work, and 70 percent of these kids are not going to finish college.

Chairman KILDEE. Dr. Sewell.

Dr. SEWELL. I think what I have heard, sir, is a distinction between the national test and a diagnostic test. I think what can contribute to the educational standard is some form of diagnostic test. That cannot be accomplished by a national test. A national test, by itself, would have to incorporate items that are more generalized to different socioeconomic groups and different school experiences across the country, and the items will be so selective that that could not have that diagnostic flavor to achieve that objective, and the information that we need to know as to what to do to promote excellence in those classrooms is already known in most of these schools.

Chairman KILDEE. Yes.

Dr. ANRIG. Mr. Chairman, let me raise a concern in response particularly to the points raised by Secretary Brock. Involved with part of what he was saying is a concern that I have with this issue nationwide, and that is that, in the long run, we turn around and put all of the responsibility back on the student for whether he or she has made the mark.

I think we need to broaden our concept of accountability so that when someone is not succeeding, we take a look not just at the student but the teacher, the school, the community, and what that community has provided to the schools, and the State for its responsibility for that school. We tend to end up, when all is said and done, saying the student made it or didn't make it, and I think there ought to be greater accountability than that.

Part of the problem reflected in your question, Mr. Chairman, is, we tend in the United States to try and use one instrument to do two purposes: improve instruction and take care of accountability. ETS, for instance, is working in areas of instructional assessment. We are working in the Pittsburgh public schools on portfolio assessments in art, music, creative writing, doing some very interesting stuff, which is great for the classroom, and great for the teacher, and helps them improve instruction, but, as a former commissioner of education in Massachusetts, I would have a tough time going before the Joint Education Committee of the State legislature and reporting that we have great portfolios, because they just don't want to hear that. What they want to hear is can the kids read and at what level, and they want a fairly simple answer to that.

So I think we need to look at ways of accountability measurement and ways of instructional measurement, and I suspect that we are not going to come up with one instrument to do both of those.

Chairman KILDEE. I want to go on to the next question, but I mentioned this yesterday, too, to the other panel. I taught Latin for 10 years, and I can recall one time trying to get across the concept of sequence of tenses, when you go from the indicative to the subjunctive mood, and finding out that I really wasn't doing a very good job. I understood it, because I had spoken Latin for four years in college, but I couldn't figure out why they couldn't understand it. Finally, through testing, I found out, obviously, they did not understand sequence of tenses. So I went back to the drawing board and put together a presentation that worked. Even the slowest student was able to master the idea of sequence of tenses.

Can testing help redesign methodology? Can it help math and science teachers collectively or individually? How can we take that back and really improve the quality of teaching?

Does anyone want to respond to that?

Bill.

Mr. BROCK. Of course, the test can do that. We have learned a lot about testing, and, with respect to Dr. Anrig here, I think that is what he was saying. I think that is precisely the purpose of an evaluative system. I like the phrase "assessment system," because that is the proper description of what we are talking about, and it clearly gives the teacher the tools that they have got to have to find out why there is a problem, define the problem, and then provide the proper response to that problem. That is what a test does if it is really done well.

Chairman KILDEE. Governor.

Mr. ROMER. Mr. Chairman, just take the fourth grade on math. I think one of the most important things we need to get together on is what are the levels of achievement and knowledge that we need to shoot for in the fourth grade, and then the test ought to clearly measure to what degree we are accomplishing that.

One of the things that seems so obvious to me is that all of us can learn but we learn at different rates. It takes some of us longer to learn than others. Again, I don't want to overuse the pilot examination, but to get a private license some people take 36 hours, some 46 hours, but we may have equally good pilots; it just took somebody longer to learn.

If we had to test in the fourth grade, for example, on math and we found out that three-fourths of the class are ready now to proceed to the fifth grade, and we had this 60 days before the end of the school term, but if that school were simply organized so that we could take that one-fourth of the class and say, "Look, we have an option of dumping you into probable failure or keeping you and giving you that extra effort"—that you needed to do in terms of your Latin students—"to accomplish this before you go on," that is something that we need to incorporate into our system.

What ought to be fixed is what you need to know and be able to do. What ought to be variable is the time you take to do it. Unfortunately, much of our education is organized the other way around. What is fixed is the number of days we sit in the seat. What is variable is what you know when you get out of that class and go into the next one.

So testing can be diagnostic, and I think Gregory Anrig said something very important in his testimony, and that is, there ought to be no test given unless it has an instructional consequence.

Chairman KILDEE. Mr. Goodling.

Mr. GOODLING. I wanted to mention in my opening remarks, and I forgot it even though I had it on the top of the list, that General Schwarzkopf said a lot of interesting things during those 100 days, but on the 101st day he said something that really caught my attention. He said it was the best educated fighting force the world has ever known. And then I heard him later, and he talked on the same subject and reminded everyone that it wasn't the high tech that won the war, it was that educated force that knew how to handle the high tech. He also reminded us that we designed, developed, and produced far better than any of the others who were—he didn't say it quite that way, but that is what he was really saying—than any of the other forces that were part of it.

I want you to know that I wrote these notes before Dr. Anrig or Dr. Sewell testified, I didn't write them after they testified.

The first question that I have is, who determines what every youngster in fourth, eighth, and twelfth grade should know and master? Designing the test to measure is simple; probably anyone could do that.

Dr. ANRIG. I wouldn't quite agree with that, sir.

[Laughter.]

Mr. GOODLING. But who makes that determination? And it might be easier in math than it would be perhaps in social studies or literature. So who determines that? Whoever wants to respond.

Dr. ANRIG. Let me take a first crack at that. I was once described by a graduate advisor as stubborn like a battleship sinking with all guns blazing. So let me try on this one.

I am a little concerned about some of the suggestions that have been made for a national board to do that. With all respect, I don't know how you would be able to set up a system which would have widespread credibility and acceptability by putting it all into one board.

I think, rather than thinking about a separate board, the Congress may already have available to it a resource that could be broadened to do this kind of function that already has a lot of credibility in all of the communities that will be affected by that, and that is the National Academy of Sciences. It has standing committees in various areas. It has just created one in education. It has a very careful mechanism for selecting people, for issuing reports, for involving consensus, and I would encourage, in response to that question, Congressman Goodling, that you consider an existing resource already out there that might be a better one to turn to than trying to create a new board where the Congress and the President and others would appoint to it, and I think a lot of people would be nervous about that, and I would be among them.

Mr. BROCK. There is a different approach to it, and that is to say that if our kids have to compete with the young people of France, or Japan, or Sweden, or Brazil, that we ought to see what they know, and that global standard could be shaped within the American curriculum, within the American experience, by a responsible group that this Congress, as a responsible body, could appoint. It isn't a board that has some ultimate source of wisdom out of their own genius. What we have got to do is to understand that we are part of the world and to compete in that world we have got to have our kids as well prepared as theirs are in a range of subjects, not just math and science but in human relations skills, in functional skills that allow them to be productive as human beings. Those can be devised and derived from the global experience.

Mr. GOODLING. Mr. Secretary, are you saying that our youngsters who have the same privileges, the same benefits, the same opportunities, as youngsters have in France and Germany and Japan cannot compete with them? I don't understand.

As you know, they have testing programs that limit who can go where and do what. In Japan, as a matter of fact, they know by the time the child enters kindergarten how successful that child is going to be and what they are going to be doing the rest of their life.

I interview every year for the military academies, and I have to say that every year the young people I interview are far superior to the people I interviewed the year before.

We have a totally different system in this country, a totally different makeup, and I think where we get into trouble is when we try to compare our overall system with the best in every other country, and I don't know that we can do that.

Mr. BROCK. That is simply not accurate, Congressman, with all respect to you. Our best only equal the average in Singapore, Japan, Korea, and Taiwan.

Mr. GOODLING. On some specific tests, Mr. Secretary.

Mr. BROCK. On virtually every international——

Mr. GOODLING. Not on their ability to produce.

Mr. BROCK. Yes, Mr. Congressman.

Mr. GOODLING. Or we would be flat on our face right now and would have lost that war in 10 days rather than win it in 100.

Mr. BROCK. Let me suggest to you that the best single training program in the world is run by the U.S. Department of Defense. Those kids, first, came out of high school, second, volunteered, and, third, went through the best training program that is available to any military anywhere in the world. Now that is the way we ought to be doing it, but we are not providing that.

This great American tradition of ours that believes in equal treatment for all children is letting 50 percent of our kids drop out in most of our major urban centers.

Mr. **Goodling**. And testing is somehow going to bring those families together, make them more literate?

Mr. **Brock**. If it gave them a better understanding of what they needed and the——

Mr. **Goodling**. Lord, they know. They know what they need. We have spent millions and millions of dollars determining what they need.

Mr. **Brock**. Let me tell you of one cross-cultural study between Seoul, Korea, and Chicago, Illinois. In that study, parents in both communities were asked how their students were doing: Are your kids getting a world class education in math and science? The parents in Chicago said yes; the parents in Seoul, Korea, said absolutely not.

Mr. **Goodling**. Be very, very careful.

Mr. **Brock**. Wait a minute now. Hear me out. In Chicago, 76 percent of the parents thought they were getting a great education; in Seoul, Korea, 23 percent of the parents were satisfied with their education. The fact is that the students in the schools in Seoul were significantly better educated than the students in Chicago.

So our kids and our parents aren't even being told what the facts are. So an assessment system would help; yes, it would help.

Mr. **Goodling**. Be very, very careful when you use Chicago, and I won't steal from Charlie Hayes.

Mr. **Brock**. Well, shall we try Philadelphia?

Mr. **Goodling**. Maybe they were interviewed where the politicians are in two portions of Chicago, where most of those children go to private schools, where those politicians who are responsible for that area make very, very sure that they don't change the State reimbursement program, because they want the rest of Chicago to get only maybe 13 percent while the other areas of Chicago get much more. So be very, very careful.

All I am trying to point out is, it is so easy to compare apples and oranges, but, boy, if ever you people on the panel should learn from each other; you need each other out there if you are going to move anything.

You know, we have a powerful person in the White House who, after they set up the governors' conference with the President, I said, "Wonderful idea. You can provide leadership as governors and as President. You can't change anything. You can provide leadership. Now you had better bring the people in who can make the change." And the response was, "Then I'd have to bring the trouble makers in." And that is a tragedy, and that is what I am worried about, that we are going to go off here and we are not going to have the trouble makers involved who are out there every day on the line.

So I think we have got to be very, very careful. It is so easy to be simplistic. You take a Japanese society, and you look at that family unit, and you look at that culture. Now they will soon change; they are becoming so Americanized, you know, they will fall. But you can't compare us with that culture.

Well, I am talking in terms of, they will be spending more time in the disco than they will be spending on homework, et cetera.

Mr. **Brock**. Is that good?

Mr. **Goodling**. Pardon?

Mr. **Brock**. Is that good?

Mr. **Goodling**. Is that good?

Mr. **Brock**. Yes. Is that what we want? Do we want them to emulate us rather than our doing better?

Mr. **Goodling**. Well, you know, the last time I visited, they were trying to make their school system more like ours. I said, "That's interesting. Ours used to be like yours, and we are trying to make ours like yours now, and you are trying to make yours like ours."

But let me ask you, Mr. Secretary, will knowing more precisely what each child knows drive teachers to teach better, drive parents to parent better, drive students to try harder? Will it help us attract the brightest and best to the profession? Will it cause parents to work and stay together to provide the care and nurturing of the child, and working with the child and the school? Will it change teacher training institutions? If it will do all those things, then I am all for it.

Mr. **Brock**. Congressman, I have great respect for you and for the question you are asking; it is the right question. There is no silver bullet in American education. Testing will not achieve all those things. It will give us one of the tools that we need to achieve those things. If parents and teachers are given more information, they will act on that information.

I think we make awfully facile arguments that we are a multicultural society. I consider that a strength, not a weakness. But when people say something like a test is to be used to screen out, dear Lord, what are we doing now? Twenty-eight percent of our kids who go to high school never finish, almost one out of three. If they live in Chicago, or Philadelphia, or New York, or Buffalo, or Los Angeles, or any of 15 other major cities, 50 percent of them don't finish, not just 28, 50 percent. The minority kids of this country are underserved. This rich American tradition forgets people when they are poor, and I don't understand.

I had an article, Mr. Chairman, an op-ed in your paper in Michigan, in which I said maybe the most important thing we can do is to challenge the constitutionality of public education, because it is not serving all of our kids and giving them an adequate education.

Testing isn't the only answer, assessment is not the only answer, but it is a tool, and if you don't know what you are doing, how in the world do you correct it? You have got to start with some knowledge base, and that is what I think we are suggesting here, not a test at the end of the process, a test throughout the process so that they can adjust themselves and learn from that, as you suggested, Mr. Chairman.

Mr. **Goodling**. Well, it was mentioned, of course, and I think Dr. Sewell wants to talk on that. We have done all this testing, we know where the weaknesses are, and, as Dr. Sewell pointed out, then they said they were going to send us some money to do something about it. They didn't send us any money to do anything about it in the city of Philadelphia, for instance.

Dr. Sewell.

Dr. **Sewell**. I think, sir, the evidence is quite clear. We know how some of the schools are functioning. We know the problems of the schools. We know that in some of the schools a teacher-student ratio of 35 to one isn't going to work for those students. We know that there is a lack of resources. The parents know that information. There has been an abundance of testing in those schools that have clearly pointed to the problems. I think the solution is something other than a national test. More information will not provide the remedial help that is necessary.

Mr. **Goodling**. I would close just by saying that we decided up here, because we heard a comment made in the testimony, and we decided that multiple correct answers will probably take care of the multiple differences of students in a national examination.

Chairman **Kildee**. Governor, did you want to add something before we go to the next questioner?

Mr. **Romer**. Well, Congressman, there are schools where we obviously know and there are places we know we are failing and we need not to solve it by more tests but by better instruction and more attention. But let me suggest that in Colorado we are losing it, I think, along with 49 other States in terms of our comparative advantage of skill levels in the world.

I think Senator Brock is correct, the facts are there, and the comparison we ought to look to is not just in Desert Storm, we ought to look at Toyota, which is expanding its capacity and General Motors which is cutting back its capacity, because that also reflects skill levels, skill levels of a people.

In terms of our work force and our skill levels and our ability to produce, we have lost substantial advantage in the last 20 years. The problem is this, in my judgment. This Nation has had it good since World War II. We have had it so good we think that to continue as usual is going to continue to have it good for us, and it isn't going to be that way. We need to recommit and redouble our effort on education in this Nation.

I think your first question was, is it going to help a student and a parent to change if they know where they are? I think it will. The problem with too many classrooms in America is, people believe that the performance they are giving is going to give them the good life, and it isn't true, and I think one of the most important messages we need to bring to the students of America and to the parents of America is that what you are doing now is going to lead to failure in competition in that global marketplace, and if you really know that, we think you will change your behavior.

I have strong confidence in the American public changing behavior if they are given adequate information, but I don't think we are giving them adequate information in education, and the issue is not tests so much as it is levels of achievement. I don't think we understand the levels of achievement we need to reach for, or, if we did, we would certainly have a different commitment to education than I see now in this country.

Mr. **Goodling**. I would just take issue—and then I am finished—with one of your statements. I don't think we have had it good since World War II. I think the decline of this Nation began at the end of World War II with the breakup of the American family, and I am not so sure that all the reforms we talk about are going to have a great effect unless we can find some other way to put Humpty Dumpty back together again.

Chairman **Kildee**. Major Owens.

Mr. **Owens**. Mr. Chairman, I want to point out that this has been one of the most informative hearings of this kind I have ever participated in. The participants also speak with great authority, and I hope that the record of this hearing can be made available very rapidly. We find it very useful in my subcommittee.

I would hope that we could understand that there are at least three agendas, and probably many more, for education. One is the academic and pedagogical instructional agenda, one is the social agenda, and another is the political agenda, which probably sits atop all that, and, whereas all three are very important, we can't have meaningful discussion unless we limit ourselves to at least one at a time, and I think what we should be addressing here is the academic, pedagogical, instructional agenda. If we can get that straightened out, maybe it will be helpful in some of those other areas; I am sure it will.

Even youngsters in inner city schools, like most of the schools in my district in New York, would greatly benefit from more accountability, and more structure, and teachers who know more about what they are doing because they have a set of standards and there is accountability in the form of assessments that are more meaningful. I don't know what those are, although you say, Governor Romer, if you have the right test, it can be an invaluable instructional tool. I hope that as you develop the right test and that set of standards, not only do you have a cluster around that set of standards that people can tie into and have different ways in which they deal with that set of standards, but I hope you have a cluster of standards, but testing for and educating for what?

We are not always educating for academic achievement, to produce scholars, and I think that is what our public school systems are geared toward. Whereas math and science is emphasized, and that is a great universal standard, global standard, and you can measure that easily from one country to another, it may not be as important as we have made it out to be. Leadership skills—I hope that somewhere there will be a standard of education for leadership. You know, some kids who have leadership abilities don't do math and science that well necessarily, although math and science can teach them to concentrate and the kind of discipline that will always be useful, whether they use it later on or not.

But I would hope that those who have certain kinds of abilities are not thrown aside, but those who can perform well, sing well—you know, we have just thrown out all of the art instruction and the music instruction out of New York City schools because they say it is not necessary, it is not vital. So without the motivation there, a whole set of kids are in more trouble than ever before.

We are forgetting that one of the most effective political leaders of this century in the United States was an actor. I didn't agree with 90 percent of his program, but I recognize his effectiveness as a political leader; all of us do as politicians. Ronald Reagan was an actor. Harry Truman was a small town lawyer. He was a lawyer, and there are a lot of lawyers around in Congress, but most of them are not lawyers who edited law journals and the Harvard Law Review. The leadership does not seem to come out exactly the way colleges and academic programs are structured. There are great politicians who happen to be scholars, but most politicians and political leaders who are very good leaders are not scholars.

So let's have a set of standards where we have a flexible set of standards, and we educate them, and know that the kid who has leadership ability may not do well in math. In fact, I understand most high school principals in New York City are ex-gym teachers. That may be bad; it may be good.

A lot of the physical education majors that we used to laugh about and joke about when I was in college are making very high salaries in business now as salesmen and business people. So they are obviously able to succeed in very difficult areas.

So I would hope that as we set this academic agenda—instructional, pedagogical agenda—we will understand the kind of flexibility where our system is probably ahead of the others in the world as being more flexible, but that ought to be, as we standardize, not lost, that we educate people for a number of different careers.

The other thing is, I would say to Dr. Sewell, on the social agenda, the Committee on Economic Development recently issued a report where they said that the social agenda in the schools is as important as the academic agenda, because youngsters are seeking learning; they are not motivated because they are held back by the fact that they have so many different problems. We must address that social agenda, but I don't think we should put down the academic agenda and the accountability, movement toward accountability. That is necessary, and we can use it even more to justify attention to that social agenda.

If kids test badly because of all these adverse conditions that they live under, let us deal with more of those adverse conditions, and I know that leads us to the political agenda, and you pointed out the fact that we are just not getting the kinds of resources and funds in the areas where poor kids live, which may lead to Senator Brock's suggestion that we sue for the lack of constitutionality, that our present educational system is unconstitutional because it is so unequal and so unfair.

So I would hope that a high level of professionalism could be brought to the whole discussion and maintained. One problem with education is, it always bogs down in these very emotional reactions and responses and we just try to cover all the territory at once, but you have done a great service by focusing on one aspect of it, and I hope we can straighten out the academic, instructional, pedagogical agenda, and I wish you luck, Governor Romer, in your labor negotiations in Denver.

Mr. **Romer**. Yes.

Could I reply?

Chairman **Kildee**. Yes, Governor.

Mr. **Romer**. It is a very good comment, but, you see, if we are looking at fourth grade math, we ought to be thinking both of those who are going to be engineers and those who are going to be technicians out on the assembly line, because a technician needs to know how to problem-solve, a technician needs to organize information and say, "This is solution one compared to solution two." Math ought to teach one to do that.

Mr. **Owens**. Problem solving is really what we are trying to get at. Math is one way to do it.

Mr. **Romer**. Fourth grade math, the test, the assessment instrument, and the curricula that relate to it ought to be teaching us to compute, but also ought to be teaching us what we would call higher-order thinking skills, which are the problem-solve, and I tell you, the factory worker needs that, also, in addition to the person who is going to be an engineer.

When you get to the 11th grade or the 10th grade in high school, one may say, "I'm going to take calculus and take a particular vocational course." Another may say, "No, I'm going to do something else and not take that course." But the critical question that I want to pose is this: I am the parent of seven children. I need to know when the youngster is in the fourth grade, is that child ill educationally or well educationally?

If that school and that teacher can't give me an accurate report on that, they have failed me, and if I, as a parent, don't take that information and, Congressman Goodling, I agree with you, the family has to take more responsibility, if we don't take that information, we have failed as a family. But I tell you, as a society, we fail unless we get a system in place that rings that bell at the fourth grade and says, "We are going to give you an accurate standard to which you reach. We are going to give you a way to test where you are and we are going to give you some remedial action to change the education of that youngster."

That is what we are after.

Chairman **Kildee**. Mr. Petri.

Mr. **Petri**. Thank you.

I learned from the morning paper why my colleague, Major Owens, has greater respect for actors than perhaps you once did since you have a leading actor in your family who is doing very well. He may end up being President some day with that acting background. That has probably softened you up on at least this one old actor's retired——

Mr. **Owens**. He hates politics, he will never make it.

Mr. **Petri**. Okay.

[Laughter.]

Mr. **Petri**. Sitting here, I guess I have a couple of questions. One is of either the Governor or Secretary. Is the assumption that we don't have adequate or good testing for tracking purposes, or is it that people aren't acting effectively on the information that is being produced by the testing that is now being done?

I understand it is being done in 38 States. Are there any where it is being done adequately or are they all failures? Is testing the magic key or better testing, or is it acting on information being produced by tests we have already?

Mr. **Romer**. Could I take a crack first, and then Mr. Brock?

It is both. We do not have adequate tests. Too many of our tests are norm-based tests, the standardized norm-based test that compare ourselves to each other and we are not adequate. We are not cluing ourselves to a worldwide standard. So part of it is the test is normative and it is on a bell curve, rather than criterion-based, which is one I illustrated for—this is what you need to know and be able to do.

Secondly, the tests themselves are sometimes inadequate. Multiple choice tests are always based upon the assumption that there is a right answer and somebody else has already posed the question. If you train a generation of scientists on that, you know, you are going to have a bad group of scientists because the challenge in science is to ask the right question and to understand that there may not be one correct answer. But it is not just the test, it is a problem of we are not acting upon the information that is there now and we need to have something that brings it more vividly to our minds.

Mr. **Brock**. Just a little elaboration. We test pretty well for people who want to pursue an academic career. We don't test at all for anybody else. We really don't know how our kids are doing in relation to what they have to be able to do in the early grades, and that is what I think Governor Romer was talking about.

He talks about the fourth grade, but at any stage in that first six or seven years of school, those are the years that you are going to make it or break it and if we don't know at that stage how we are doing against a standard that, in my judgment, has to be globally derived, we don't know how to change what we are doing or how to adapt or how to respond.

We underestimate our teachers. They are hungry for this kind of information; they are desperate for it. They are eminently capable of doing what is necessary if they are given the tools, but what we do is we say, these are the tools you have to do if you want to get somebody to go to the University of Tennessee.

Mr. **Petri**. The nice thing about education is that everyone who is an adult is an expert because they have all experienced it from the point of view of being a student. All I can remember is that I was tested in public schools in Fond du Lac, Wisconsin, from morning kindergarten on, and I thought they were fun. I thought they were games when I was very young. There were hearing tests and eye tests and recognition tests and 101 different kinds of tests all the way through. Are we talking about what is already being done? You, sir, are in the business and maybe you could share with us your thoughts on whether the problem is really figuring out better ways of assessing or acting on the information more effectively?

Dr. **Anrig**. Thank you, Congressman. It won't surprise you that I think we know how to make good tests. The issue is not whether we know how to make a good test, but do we know what we want to make the test for, and that is why I believe, and each of the speakers today has said, first you have to decide what are you trying to accomplish? What is it you want a youngster to know and be able to do? Then you design a test to measure that.

As Congressman Goodling said, anyone can—I forget the expression used, but I don't think I would want to put it back on the record, but we know how to measure knowledge, but we can't and shouldn't be the ones who define what that knowledge is. That ought to be designed through a very broad-based consensus process.

Mr. **Brock**. Tom, can I just add one point. In today's workplace, there is a radical difference in how we organize work and the really good companies, Motorola, Xerox, firms that are way ahead of the curve, that are doing what other firms are doing around the world, and they are becoming very competitive. If they are not reorganizing work, by the way, they are not staying competitive and we are trying to compete on the basis of wages, but those companies that compete on the basis of skills are finding out that they have to focus on quality and they have to focus on teamwork and they have to focus on horizontal organization, which means workers individually have to carry more responsibility.

Now, we are not teaching those kinds of skills, nor are we measuring those kinds of skills and you can start teaching team process as part of a math session in the fourth grade. But it requires somebody to tell the teacher, this is what we have to have to be globally competitive and we are not telling them that now. That is the point.

Chairman KILDEE. Mrs. Unsoeld.

Mrs. UNSOELD. Thank you, Mr. Chairman.

I came in here sort of open to learn and I found myself fast becoming a fanatic.

[Laughter.]

Mrs. UNSOELD. Governor, I share your fervor when you spoke that when the bell rings at the fourth grade, we need to do something, but I want to submit, "the bell rang a heck of a lot earlier." We have used some Persian Gulf examples and the one I would use is that when the Patriot was begun to be developed, we had in the United States 70 percent of the world manufacturing of consumer technology and we are now down to 5 percent, so I think we all agree, Mr. Brock, our nation's economic future in international competitiveness depends on improving our country's educational system. We have a problem.

But what I hear is that more accountability is going to cause correction to be made. Again, Governor Romer, you said that knowledge of where they are on that graph ought to motivate them and, Secretary Brock, you said let the student know where he is and then some correction can be made.

But both of those statements assume that there is motivation and they assume that education is beginning at about the fifth year of life, and it isn't. It is starting way earlier and we know that the kids coming into kindergarten are much less prepared to succeed today than they were 10 years ago and 20 years ago, so the bell has already rung and we have been warned and we are not making that correction yet.

Probably by age two or three, the motivation for learning is pretty well established and any great quantum leap forward you may make after that is overcoming an extraordinary handicap. We have to aim it at that earlier age.

We know that the investment there is going to pay off. We know that when we find pregnant teenage girls and we get them into a special kind of high school class, that the offspring produced there are going to perform better when they become school-age. We know we can make that investment, so the bell has rung, and gentlemen, it is not assessing it after we get to school. Let's learn what we already have and let's concentrate that fervor that you have to making sure that we are paying for those early childhood education investments, we in this country. If we fail to do this, our nation will not make it in this next century.

Mr. BROCK. Mrs. Unsoeld, if I may just jump ahead of the Governor for a second, neither one of us believes that testing or assessment is anything more than a very partial part of a solution. For myself, and I think I would probably speak for almost everybody here, if we don't start dealing with the front-end problem, we are never going to solve it with the back-end solution. We have to deal with pre-natal nutrition, post-natal nutrition. We have to deal with early childhood education. All of those issues are fundamental to correcting the problem, but no single answer is going to deal with this problem and what we are saying is that today we are talking about one aspect of it that really has to be addressed. We do have to know where we are.

The teacher has to know where they are. They have to know what they are being asked to do and we are not telling them.

Mrs. **Unsoeld**. I submit that the teacher knows that without a national assessment, but the teacher is so having to concentrate on making up for the failure, as Mr. Goodling put it, of our families having become dysfunctional, that they are having to try to remedy for what was lost in those early years.

Mr. **Romer**. Could I make a comment?

Mrs. **Unsoeld**. Yes.

Mr. **Romer**. You are absolutely correct. The panel which I chair has the obligation in the next two months to figure out what are the five or six most critical questions to ask each governor to report on in terms of goal number 1, every child shall be ready to enter school by the year 2000. The posing of those questions, Mrs. Unsoeld, are really critical and I understand that.

But let me come back to the Federal establishment here. What we have talked about this morning is a failure of research and development in assessment in this nation. We have been talking about it for two hours. We have not done the kind of research and development in terms of the standards or the assessment instruments or how we use them.

Now, let me ask you, since we have been talking about the Patriot missile, go back and compare the amount of dollars that have been spent in the last 20 years by the Federal Government on research and development on weaponry as compared to research and development upon what are we doing with assessment and goal setting in our educational system?

If you want a classic failure of allocation of resources to the right area, look at your own budget in terms of research and development as it relates to education.

Mrs. **Unsoeld**. I join you in your fervor and let us combine it and try to make that repriortization in our funding.

Chairman **Kildee**. Mr. Hayes.

Mr. **Hayes**. Thank you, Mr. Chairman.

I am really caught up in this testimony very much and I am very much a stickler to time. I realize the difference between the green light and the red light is only five minutes and I have to observe that.

But I just wish we had an opportunity to have some of this testimony heard by the entire 435 member body which I am a part of. Maybe we would get serious about reforming our educational system.

When you talk about making it our priority, it sure isn't. We just spent $30 billion yesterday as a part of the bail-out program and I understand that on track is another $50 billion, I guess, by September. Yet, when you look at the amount of resources we are willing to invest in our educational system, it seems to be such a struggle, you know.

I happen to represent, as my good friend Goodling has already indicated, part of the City of Chicago, which happens to be the poorest, a school district that the amount of money spent per student is much less than the amount of money spent on the student that surrounds Chicago in the suburban areas.

The main reason for that is the tax base that is used in order to fund the school system. Illinois, as a State, doesn't have a good record as to how it disperses money to the various school districts. What they use is a property tax base as a means of doing it. There is no property—if the value of property is down, the amount of money you get for the schools is down.

We got, I guess, the highest dropout ratio of any part of the State of Illinois. It runs about 50 percent in many of the high schools in my district between the kid who starts the high school as a freshman and one who stays long enough to get a diploma.

Testing is not going to solve that problem. I agree with much of what you said about we ought to have assessments at different levels as to how a student is doing and the check with the faculty, the teachers and the parents. Parental involvement is very important.

I do want to ask a question of you, Senator Brock. I have reached the point where I have almost thought that we are willing to write off a lot of kids because of the economic status. You don't care whether they get an education or not, forget them, you know, even as we talk about choice, and this is a big issue. What is going to happen to that poor kid whose parents can't choose what school they go to. That is the school that is going to get the money. Part of the criterion for getting in that particular school is going to be your grade level or your achievement.

What is the impact—this is my question—for students who do not achieve the standards that you would establish by age 16? Will we provide other opportunities for them to achieve or would we begin to track students at an early age in the areas that they perform best? How do you deal with that kind of problem?

Mr. **Brock**. What we suggested, Congressman, in the Commission on Skills of the American Work Force, was that we, as a society, accept that as our responsibility, and Dr. Anrig was right, we are not talking about putting all this burden on the students. It is our burden to get them up to a level of productivity that is globally competitive.

We said that we ought to try to commit to get everybody up there by 16, but as Governor Romer has said, not everybody moves at the same pace. Some move slower, some move faster.

If they make it by age 16, then we have opened up a whole set of choices for them that they can move to. If they haven't—and let's admit that there are some kids that can't make it in public school, for whatever reason, family, Congressman Goodling is absolutely right, there is a fundamental family problem in this, but whatever the causal factor, there are some kids that aren't going to make it in the public school setting.

So what we said is, look, that doesn't relieve us of our obligation. We need them as participants in this society. We can't constantly carry 30 percent of the public of this country on our backs. Everybody has to have an opportunity to be productive, so let's set up alternative learning centers and let me tell you about one, because we already know that this thing works.

In San Diego, there is a guy named Tony Trujillo, that runs a public school. It is the biggest district in the United States. In that district, they have decided that the money follows the child and if the child drops out of public school, they have an alternative program for them that they put them back into some form of education. The fact is it is in the same school community because they think kids like to stay with their peers, but it is a different process entirely. It is radically different.

Do you know what the average daily attendance of dropouts is, these 2000 dropouts, is 97 percent. That is higher than it is at the Naval Academy. Do you know that they are out-performing the students in the regular classes? Do you know they are learning faster and more than they do in the regular classes? We know what works. We are not doing it. That is the point. We know we can do it. We know we have to do it. It is insane to think we can tolerate 30 percent of the kids in this country falling off by the side of the road. This country can't compete with anybody if we don't educate those kids and we have to accept the responsibility of doing it.

We can't do that, though, unless we know they have a chance of falling off the track and that is where the assessment comes in. It is just a limited part of the answer, but it is part of it.

Mr. **Hayes**. All right.

Chairman **Kildee**. Dr. Sewell, you wanted to add something?

Mr. **Hayes**. He is itching to do so.

Dr. **Sewell**. I think the notion of an alternative learning center is a very desirable one if we can identify some of the strategies to make it a valuable one for the kids, but that is one of the areas that we have to be very cautious about when we think in terms of assessment because the history of testing in America has been used for tracking, sorting and putting individuals who are capable of the highest intellectual performance into those tracks. We think in terms of a State that is testing kids in kindergarten to determine who is ready for first grade.

If we start that tracking system and use standardized tests to accomplish that goal, we will start the weeding process very early and undesirably so.

Mr. **Hayes**. Some of us call it creaming.

Thank you. My time is already gone.

Chairman **Kildee**. Mr. Roemer.

Mr. **Roemer**. Thank you, Mr. Chairman.

I would just like to thank the distinguished panel. I planned to leave about an hour ago to attend another committee hearing, but this has been so fascinating that I chose to stay.

I also hope, Governor, that when you issue the report card from the National Governors Association you give the same attention to the Congress that you do to the White House because we must work closely together on this problem. Congress is very, very concerned about this issue. Maybe you could come up and present the report card to us and work with the chairman on an idea of that sort.

Coming from Indiana, I would just like to say that the education problem is the most critical one that we face. We all like to make the analogies to the Persian Gulf War or World War II. But education is the essence of our society and our children's future. In a rural county in my district, the dropout rate is 25 percent. The Director of Correctional Facilities said to me the other day in Indianapolis, "The biggest indicator of prison space in the next 20 years is the number of at-risk kids in second grade."

That is frightening thought for us as a community and as a country. Yesterday I heard testimony from another distinguished panel that our facilities in colleges are falling apart. In 1991 we are spending in constant dollars for instrumentation and facilities modernization half as much as we did in 1966. We have tremendous problems which need America's attention. We must work together to solve this problem.

My question is, given the limited resources we have, and the limited capital, where would you place national testing in terms of the scope of the problem that we face?

Mr. ROMER. Could I give my own personal answer? I have, in the back of my mind, about 17 things I think we need to do to reform education. This is one of the 17, so it is not the full answer.

Mr. ROEMER. Where does it fall, Governor?

Mr. ROMER. The matter of setting standards falls right at the top. If we do not determine what it is we are trying to do first, all the other 16 reform elements in my mind don't fit. I would say not testing, per se, but arriving at the levels of achievement, a national consensus of what it is a youngster needs to know and be able to do, I believe ought to be the number one item in terms of the timing in which we approach it.

Now, it isn't—teaching and what happens in the classroom is finally what is really the most effective thing that makes a change in education, but I don't think that we are clear yet about what are the levels of achievement that we need to reach for and I would just say, it is one of the very top items, but it is only one of a long list of educational reforms that we need to accomplish.

Mr. BROCK. The problem is, if you do any one of these things, you may do something that is counterproductive because they are hooked together. I mean, people talk about choice. If you have choice without school base management, you are probably making a mistake. If you try to have school base management without giving the teachers the assessment tools, you are probably making a mistake.

These things link together in a coherent whole. The reform of public education has to go across the whole range of subject matters and it isn't a matter of resources, Congressman. We—Lord knows, if you think education is expensive, try ignorance and just see what this country is going to do if we try to compete on the basis of lower wages all the time.

We don't have that luxury and it isn't a matter of spending money. All the research that is being done in assessment is being done in the private sector. Where in the world is the Federal Government? We don't do any R&D and that is one role that this government ought to be able to do at a very low cost and high ratio of yield.

Mr. Owens was touching the right point earlier. He was right on the money when he said, we have to teach reasoning skills, logical skills and that means different things to different kids. But we are not saying a test is going to answer. We are saying that if you compile a series of evaluations over a period of time, you are giving the benchmarks to that student and to that teacher and to that parent and to that principal that they have to have to move in the right direction affirmatively and if you don't have those benchmarks, if you don't have those milestones, we don't know where we are going.

Today, we just don't have them.

Dr. **Anrig**. Let me, if I may, just add one further point. I want to strongly endorse the points that have just been made by the Governor and Secretary Brock. You can't look at assessment as something separate unto itself. If you do and you act just on that one piece, you are apt to foul up the whole deal.

What we have to do, and I agree, the absolute first step must be clear up the standards, then have some way of measuring how we are doing on that and then, Congressman Petri, be sure the kid is getting the help and the instruction that makes it possible for him or her to succeed with those standards.

If you do them through the latter part, don't even bother doing the first two.

Mr. **Roemer**. Mr. Chairman, may I follow up for just a moment?

Chairman **Kildee**. Certainly.

Mr. **Roemer**. Secretary Brock, since you have a great deal of expertise with the America's Choice study and the Work Force 2000 study, what specific recommendations would you make in vocational education? How should we involve parents and the business community in this aspect of education as well?

Mr. **Brock**. Let me say that I almost feel like we ought never to use the word vocational education again. All education ought to be training you to be productive as a human being. That is sort of vocational in the purest sense of the word.

What we are trying to do with the Secretary's Commission on Necessary Skills, the SCANS Commission that Elizabeth Dole asked me to chair, is to see if we can work with labor and with business to describe what we call functional skills that are necessary to be employable and functional as a human being, and with a high growth path.

So we are saying if you are going to teach math—and math is an enabling skill; you do have to be able to make change in restaurants; you have to be able to count a tip or pay a taxi; there are certain basic things. But shouldn't we be teaching math in that context as a functioning skill, not just as a numeric column of numbers?

We do the worst job of an industrial nation in the Free World of communicating to students the relationship between what they work on in school and what they are going to work on in life. The school-to-work relationship is not established in this country, and that is part of the problem.

So your question on vocational may be precisely the right question. Can't we reorient our education so that we are creating a sense on the part of the student as to why they are learning and how it would work for them, what it would mean to them? Can't we teach them functional skills as a part of that math, as a part of that science, as a part of that English training? If you do that, then you are beginning to create the linkage that really does give a sense of motivation and logic as to why they are there.

By the time the kid reaches 16, if they haven't dropped out already, they are bored silly for the next two years, 90 percent of them. Why? Because they don't know why they are there, and, frankly, business doesn't either; there are not five businesses out of a thousand that ask for a high school transcript. What is the message? "It doesn't mean anything. All we want to know is, will you be at work on time?"

Mr. **Roemer**. My question is, how do we get those businesses involved in the schools and in that kind of training? Business representatives come to me saying they want to be involved.

Mr. **Brock**. If you can get businesses to help us describe what those functional skills are, then you measure those skills and you come up with a series of what we call merit badges or certificates that a business can take and evaluate, and all of a sudden you have created that linkage where the business has a vested interest in the product of the school system, and you are creating a sense where the business has an investment in trying to make it better. If you don't create that linkage, you are not giving the system the pull that it has. We have got the push coming from underneath, from the kids. We need the business community to pull in the right direction. That is sort of the priority we are trying to establish in the SCANS work. Can we motivate the business community to tell teachers for the first time what they want? If they do, we can describe that, then we will have to learn to teach it, then we will have to learn to measure it, and then, all of a sudden, we have got a system that a young person can understand and a business can understand, and it begins to compute for them.

Chairman **Kildee**. Governor Romer.

Mr. **Romer**. I know we are getting near the end of our time, and we have just spent two hours describing the problem. There is a way that we can go at this jointly, and I would like to take a minute or two to just describe that.

If we get our minds together—Congress, the executive branch, the governors of the Nation, and all the other educational interests that have something at stake—we can go at this, and we say, "Hey, what we need to do first of all is to arrive at levels of achievement in the major subject matter areas." That can be done. It can be done on a collaborative, a consensual basis.

On math, you start with those who know it best, like the national counselors, the teachers of math. You take what has been done to date at the national level, and you distribute it to the Nation and say, "How do you feel about that?" You get their reaction. Then you come back, and you arrive at a standard, and it can be one which can be changed from time to time, and then, systematically, you begin to do that with other subject matter areas.

We arrive in this Nation at what specific levels of achievement are needed, and then that is a portion that is done. Then we turn to the question of assessment, and we say, "Look, you in the States have a lot of resources already you are spending there. Let's help you combine so you can get more bang for the buck."

Congress could come to the table and say, "Look, we will put some R&D money on the table to help do this, but rather than us do it at the Federal level, that's wrong, we will do it with you." You could put some matching grants on the table, and you could design a program in which you could aim at assessment that doesn't just hold a report card up, but assessment, as many of you have said, changes the system.

Then we can go from there, from standards, to assessment, to looking at, all right, what are the other 16 educational reforms? and then let's get a partnership approach to that from the Federal and State level.

But as a nation—and this is what the educational goals are all about—we ought to say, if we really believe in that, let's put it right at the top of our priorities, let's get a program that is 10 years in length, and let's start down that path. But we ought not just sit here this morning and say, "Hey, we can't get there from here." We can, but I think the way to start is what we have been covering this morning, and that is to start with some consensual understanding of what it is we are trying to achieve, a level of understanding of what a youngster should know and be able to do in various courses at various age levels.

Mr. **Rormer**. Governor, I think that is a great idea. In my home State, they are starting to lay off teachers. We all get up in front of our microphones and say how important education is, but we must come up with some ideas for these partnerships and teamwork between our States. I hope we can do that before the report card.

Chairman **Kildee**. Mr. Petri.

Mr. **Petri**. I just have one quick follow-up question.

You focused on dropouts and people who don't meet the standards. What about the people who very easily meet them, who aren't challenged enough? Is this a national problem, that we are underperforming because we are not utilizing resources out there of people who, with no problem, do the test in half the time, and they are ignored somewhat by teachers because there are other kids who have problems? They get bored. Are we losing as a country in competing because of not demanding enough? And, if we have national standards, isn't there some danger that they may be set too low because we will not want to embarrass people, that we are trying to figure out a kind of reasonable thing where most people will be sufficiently challenged, and therefore we will set it there so that most people can be engaged, and not set it at the level that will challenge people in the top 20 percent or so?

Mr. **Anrig**. Congressman, you raise a very key point. We should not set standards in such a way that there is a single fixed point that we expect all kids to reach, because that is exactly what you are going to get if that is what you set. What we ought to have are striving standards, maybe levels of them, if you will, so that a youngster will feel that he or she is making progress towards a level but also can go well beyond it.

The National Assessment Governing Board has got three levels in the standards that they are working on now—basic, proficient, and advanced—so that youngsters could have some feeling of movement and schools could have some feeling of movement, and not just a single, fixed point.

I think we have got a mentality in the United States that we want to put on the front page of every newspaper one number that covers everything, and that is part of our problem.

Mr. **Brock**. One of the things that we have wrestled with in the SCANS Commission is precisely this question. The one thing we are desperately trying to avoid, and being very sure we are going to avoid, is a pass/fail kind of approach. That is not what we are about. We are talking five levels, but the same kind of a concept of, nobody fails, nobody exceeds, but there is a range within which we do expect, and I think we can use the system to motivate people at every level if we do it right.

Chairman **Kildee**. Mr. Goodling, for a concluding remark.

Mr. **Goodling**. Just two.

Some of the finest research that has helped education and also the medical profession has come from the military research, and so that wasn't all bad.

If I sounded like I was getting very excited about this issue, about reform, I am. I get excited because my concern is this. In the middle of this panel we have two very successful, very outstanding gentlemen who should be driving the people who are flanking them. My fear is, as we go through this reform process, that two gentlemen times 100 gentlemen just like you—and women, very successful, very outstanding—will try to drive this road alone, and we will not have the gentlemen flanking you in that automobile, or the four that we had yesterday testify. That is my greatest fear about the whole reform movement.

You both are in a position, because of who you are and the successes you have, to drive. Perhaps the two who flank you aren't quite in that kind of position when trying to drive the public. But my fear is that those like you may try to drive this road alone, and from a lot of the reform movement that I have seen, that has taken place. That is doomed, in my estimation, to failure.

If you can drive the two who are flanking you, and if they can drive with you, then I think we can come up with the kind of reform in this country that we positively must have.

Mr. **Romer**. Congressman, that is very good, and you and I have talked privately about this. I want you to know, I am responding to some of your thoughts.

The goals panel in the month of April is going to go to eight places in this Nation and lay out where our thinking is to date and the expertise that we have heard to date and ask the gentlemen in education who are on site, doing it, to share with us their criticism and their input, and I think that all the way along in this goals panel project we really do need to have it very inclusive, very inclusive of the people who have a stake in it, and particularly those who are on the front line, the teachers. We really need to make it inclusive, and so I take your comments seriously and have incorporated it in the life of the panel.

Mr. **Brock**. I think it is a super point, and I couldn't agree more, but I think you overestimate the two of us. I would just like to be in the education car; I don't have to drive it; I just want to contribute somehow to moving this thing, because I am truly, truly worried about this country right now. I don't believe we can compete with an uneducated work force, I really don't, and I don't think we are doing an adequate job by our children.

Mr. **Hayes**. Could I raise just one issue? The Senator caused me to have to raise this final question with your comments now.

When do you plan to submit your final SCANS Commission report?

Mr. **Brock**. We are going to have our first phase of the report, hopefully, in May, and then we will try to refine that over the following six to eight months after that, and have a completed product by early next year.

Congressman, it is our intention to set down the basic benchmark of this effort within the next two months.

Mr. **Hayes**. That is some pretty fast tracking.

Mr. **Brock**. Well, it is faster than I thought possible, but the surprising thing to me has been just how active people have been in participating in this effort. We have had really astonishing support from labor, from business, from educators, and, frankly, we are moving faster than I thought we could.

But understand that what we do in May is going to be very tentative. We are going to put it out there and say, "Kick our teeth in," and we will get a lot of people who will try to do that. We will then reshape it and reform it and try to adapt to those criticisms and see if we can make this thing something that is workable. I hope it will be, but it is a big task.

Chairman **Kildee**. I want to thank the panel. I am not exaggerating in the least when I say that we really have assembled one of the finest panels that I have had the privilege of listening to in my political career. We have had concurring and contrasting views. I think Bill Goodling, as he always does, gives good advice; I think it was very good advice and you all accepted it very, very well.

What I intend to do—because I think this hearing has been very, very helpful—I intend to put the printing of this record on a very fast track so that the educators, and elected officials, and the public will have the benefit of your very insightful testimony, and I personally thank all of you for being very, very helpful to the Nation. Thank you very much.

Mr. **Brock**. Thank you very much.

Chairman **Kildee**. Our next panel will consist of Dr. Saul Cooperman, president, Educate America, Morristown, New Jersey; Dr. Karl V. Hertz, superintendent, Mequon-Thiensville School District, Mequon, Wisconsin; Dr. Monty Neill, associate director, FairTest, Cambridge, Massachusetts; Dr. E.W. Kelley, professor of American Government, Cornell University, Ithaca, New York; and, an old friend, Gordon M. Ambach, executive director of the Council of Chief State School Officers, Washington, DC.

Dr. Cooperman, you may lead off, unless you have made some other arrangement among yourselves.

STATEMENTS OF SAUL COOPERMAN, PRESIDENT, EDUCATE AMERICA, MORRISTOWN, NEW JERSEY; KARL V. HERTZ, SUPERINTENDENT, MEQUON-THIENSVILLE SCHOOL DISTRICT, MEQUON, WISCONSIN; MONTY NEILL, ASSOCIATE DIRECTOR, FAIRTEST, CAMBRIDGE, MASSACHUSETTS; E.W. KELLEY, PROFESSOR OF AMERICAN GOVERNMENT, CORNELL UNIVERSITY, ITHACA, NEW YORK; AND GORDON M. AMBACH, EXECUTIVE DIRECTOR, COUNCIL OF CHIEF STATE SCHOOL OFFICERS, WASHINGTON, DC

Dr. COOPERMAN. Mr. Chairman and members of the committee, I am Saul Cooperman, president of Educate America. It is a nonprofit, nonpartisan organization in Morristown, New Jersey. I am here to represent former New Jersey Governor Tom Kean.

I heard two comments of the former panel and of the representatives that I want to comment on first, because I hope the baby isn't thrown out with the bath water.

First, local control—that if we have any sort of national assessment system or national exam, we are hurting our local control. I have great respect for it. I have been a teacher, principal, and superintendent for 23 years before I was Commissioner of New Jersey and my present position now, so I have respect for local ingenuity.

But I would ask the panel and I would ask everyone to just consider, when we keep saying local control, do we realize that we have a national textbook system and four or five textbook manufacturers say what is taught? The teachers are trained essentially the same way by departments of education and the deans. We have reciprocity in almost every other State. The school year is almost the same, and the vacations are the same. If I blindfolded anyone and put them in Arizona, or Alaska, Missouri, or Michigan, in the fourth grade, they would see essentially the same schools. So what I am asking you to consider is, when you consider national assessment, don't mix and match the local control issue.

The second, the deep problems in our cities. Congresswoman Unsoeld and Dr. Sewell, I thought, spoke rather eloquently. There are no easy solutions. I spent four days each week in the city of Newark, New Jersey, trying to find solutions. There are not easy answers. But because we have problems in other areas does not diminish the need for national assessment.

We believe we should have required national achievement exams for all high school seniors and all fourth graders and all eighth graders, not some, not a sample, all. Such an examination—and I will focus on the seniors—would provide a reliable accountability system to students, to their parents, to schools, and to the States for the first time in our history.

I believe we should have a core curriculum. We ought to state what children should know and determine whether they know it or not. Therefore, I am in agreement with almost everyone on the preceding panel. If kids don't know what they are supposed to know, we have got to challenge schools.

Someone asked the question, what do we learn from the exam? Well, a very simple thing we would learn from the exam. If the teachers aren't trained to teach what they ought to, we can drive the results back to the teacher colleges. We found that in New

Jersey. We gave a test and found out that in one of our colleges 50 percent of the teachers could not pass a test in their academic major; their grade point average was 3.2. The rigor was diminished, the grades were inflated, teachers were passed out, and they couldn't teach their subjects. A national examination would key that.

I want to give you a concrete example. I would like to give more, but I will give one of what a national achievement exam will do and create things for the better.

In many of our high schools, when the kids don't attend college, they know a diploma is important; they know employers are going to ask, "Do you have a diploma?" They are very savvy. The know the diploma is important, but that is it. In the last 25 years—and Ernie Boyer points this out in his book "High School,"—the general track has increased dramatically; more kids are taking general science and general math than biology and algebra, and why not? They all count for five credits. So as long as you get the credits, you get the diploma, and the employer asks for the diploma.

We have given the message to students that taking tough subjects does not count. We have got to deliver the message that working hard and achieving means something, and there will be something else. You talked about linkage, and the other gentlemen talked about linkage too. Here is real linkage. If we link accomplishment in school to the world of work, then you have got a hell of a connection. The message should be: Do well in school, and you will get a better job at better pay than if you graduated with a mediocre record.

Until now, there was no way to make these words and deeds the same, but because the President and the governors have established a consensus on goals in reading, writing, math, science, history, and geography, we have the willpower to take the next logical step and ask with seniors, and then eighth graders and fourth graders—I would do it the reverse way—we will have a state-of-the-art system that testing people—Greg Anrig has told me, and others have told me, "We know how to do that, Saul. Set the goals, set the standards, get a national consensus, create the exam; then we are operating from a basis of fact, we are operating from a basis of real data, data that we don't have before.

This does not mean top down. As everyone so eloquently stated, it can mean real inclusion. We can get the people who have the credentials, let everyone in, hold meetings across the United States, let anyone come forth, make it inclusive, great involvement, but have a national examination. Then business people will ask students, "So how did you do? What do you know?" And we at Educate America have a five-point scale; I could talk to that later, if you wish.

The National Achievement Examination has the power to change things, and it will galvanize the public, it really will.

Governor Romer was talking about, you have got to have an accurate picture; this will give an accurate picture. And Governor Romer said the man on the street must understand it. One can be simple without being simplistic; one can be direct and understood. We can give a national examination.

Right now, most people in this country pay little or no attention to international comparisons or NAEP exams. I was on NAEP for two years as the commissioner's representative, and then when it became the National Assessment Governing Board I was on for two years. I used to go home in New Jersey, and I was all fired up and very upset, and I would get teachers, and administrators, and board members, and parents together, and I would say, "We're not doing very well; we really aren't; the data is there;" and they would say, "Saul, what are you getting upset for? What are you getting upset for? Our schools are doing okay. No one knows; they don't compare what we are doing with Osaka or Munich. Why should you worry?" You see, they didn't worry, because it was a sampling of students in the United States; it wasn't brought back to their door, it wasn't brought back to their school, it was kind of antiseptic and anonymous. Because the results were based on a sampling of students, the message never came to the specific school.

With a national achievement examination, results would be known for every high school and every high school child in the country. The message of results, the message of facts, would be publicly delivered to every high school door, and facts are stubborn things. For the first time in our history, we would have a national achievement examination, a national exam, with the same questions asked in California and Connecticut. I don't know why we can't do this. Photosynthesis is the same in California as it is in Connecticut, and the quadratic equation is the same in any one of our States, it doesn't change. It should be an achievement exam, not an aptitude exam. It should be linked to the curriculum so everyone knows what is expected. The objectives—the teachers would teach in a variety of ways to meet these objectives, and the results would be known. For the first time in our history, we would have goals and accountability; results will matter, and attention will be paid.

Thank you.

[The prepared statement of Dr. Saul Cooperman follows:]

Mr. Chairman, members of the Committee, I am Saul Cooperman, President of Educate America, and I am pleased to be here this morning on behalf of former New Jersey Governor Thomas H. Kean. I thank you for this opportunity to present Educate America's proposal.

Educate America believes we should have a required National Achievement Examination for all high school seniors. Such an examination would provide a reliable accountability system to students, schools and states.

Educate America believes that we ought to have a core curriculum; we ought to state what children should know in key subject areas and determine whether they know it or not. If they don't, we should challenge the schools who don't measure up, and learn from those who do. Make the ends drive the means, make results drive the system.

Let me give you a specific example of the present system and how Educate America's idea for a National Achievement Exam will change things for the better.

In every high school, students who do not plan to attend college know that the diploma counts, but not much else matters. These students know employers will ask "do you have a diploma" so it is important to them. But, during the last twenty-five years the "general track" has increased dramatically as more and more students take general science and general math rather than Biology and Algebra. Why not, they reason, all are worth five credits.

We have given the message to students that taking tough subjects doesn't count and that hard work doesn't count. Educate America believes we've got to get our educational system back on track. We've got to deliver the message that working hard means something; results mean something. We've got to link real accomplishment in school and the world of work. The message should be, "Do well in school and you'll get a better job, at more pay, than if you graduate with a mediocre record."

Until now, there was no way to take these words and make them a reality. Now we have a way. President Bush and the nation's Governors have established national goals in reading, writing, math, science, history and geography. If we have the willpower to take the next logical step and ask what seniors must know in these areas, we can then create a state of the art examination to determine what students really know and are able to do. Then businesspeople, who ask students for their scores can have solid information upon which to reward accomplishment.

A National Achievement Examination has the power to change things by galvanizing the public. Results will be clear and easily understood. Right now people don't pay much attention to international comparisons or NAEP results. As some New Jersey educators told me a few years ago, "Saul, I know the general picture might not look too good, but our schools are okay, so why should we worry." Because the results were based on a sampling of students, the message never came to the specific school. With a National Achievement Examination, results would be known for every high school in the country. The message of results, the message of facts would be publicly delivered to every high school door. And facts are stubborn things!

For the first time in our history, we will have a National Achievement Examination. A national examination, with the same questions asked in California as in Connecticut; an achievement exam, not an aptitude test; an examination for all seniors, not a sampling of seniors. We will, for the first time in our history as a nation, have goals and accountability. Results will matter and attention will be paid. What gets measured, will get done.

EDUCATE AMERICA/page 2
SC/kk
0600

Chairman KILDEE. Thank you.

Dr. Hertz.

Dr. HERTZ. Mr. Chairman and members of the subcommittee, on behalf of the American Association of School Administrators and on my own behalf, I would like to thank you for the opportunity to testify on this most important issue.

I am superintendent of the Mequon-Thiensville School District in suburban Milwaukee, Wisconsin, and if Mr. Owens were here or if one of his staff members is here, I would like to further establish credibility with him by saying that I was an English teacher when I was in the classroom.

During the past year, I have served on the Wisconsin Governor's Commission on Schools for the 21st Century which flowed from the President's gathering of governors regarding education, and I chaired the part on assessment.

AASA would like to address three issues surrounding the development of national tests. First of all, can a national test be the major force in improving education? The short answer is no. A national test cannot be the major force in producing educational quality. First, a national test would hold students responsible for conditions they do not create nor do they control.

An associated dramatic concern connected to national testing is the inappropriate placement of disadvantaged children, especially those who come from nonstandard English speaking homes and classes that do not challenge their true talents.

Second, a national test would occur separately from the daily learning process. Because the test would be removed from the instructional setting, the results are unlikely to be important to the students themselves. Further, the Wisconsin Commission's report identified the relationship between the assessment process and the immediate and direct application of that information back into the instructional process.

Third, the notion that quality can be achieved chiefly through an end-of-the-process inspection is wrong. Quality cannot be inspected into a product; rather, quality can only be built in. Edwards Deming is all over this country and all over this world talking about that in all sorts of forums.

Fourth, developing a single national test will take a long time, and, given the need to hold down expense, the test will likely have to be multiple choice or short answer tests.

I would also suggest that the construction of a national test will be fraught with political struggles which will neutralize any potential value it might hope to have.

Fifth, a multiple choice, short answer test will almost certainly fail to accurately measure student learning, because a national test will not test what is being taught.

Since there is no national consensus about what students need to know and be able to do, no national test has any basis for choosing what to test.

The second AASA issue: Can test scores force educators to do a better job? Again, the short answer is no. Teachers and administrators whose students do not score well can be humiliated, but in the absence of clear information about the causes of failure to learn, improvements cannot be made.

AASA's third issue: Is there an alternative to a national test that can provide information about what students know and can, at the same time, be useful to both instructional processes and political leaders?

The short answer is an emphatic yes. School districts that have made great strides using information describing student learning, mastery, and have traits in common—first, such school districts began by clearly describing what students are expected to know and be able to do when they finish school. Second, the districts did a careful study of how they could best facilitate the desired outcome through structure and curriculum. Third, they spend a considerable time in teams deciding what evidence of progress and mastery would best facilitate learning and improve teaching. Finally, the school districts carefully studied what more they needed to know and be able to teach every student the desired outcomes and use information about learning to continuously improve.

There must be a consensus on what students are expected to know and be able to do, otherwise there is no basis for judging growth. And the fact is that different districts and different locations and at different points in their history have varying expectations, goals and aspirations. When a national test is given in such a setting it can quickly become a common level of minimum competency, which can cause huge numbers of students and other people to focus upon an expectation which is considerably out of sync with the condition of a given community.

During the deliberations of the Wisconsin Commission on Schools for the 21st Century the Assessment Committee heard over and over again that norm reference tests were not used to improve the instruction of children. When our district receives the results of national testing or is doing assessment of writing or individual testing by teachers in the schools, we take the results into small groups of teachers and building administrators and plan exactly what the educational strategy will be to effectively use that information to better instruct the children.

We feel that an immediate use—and I would like to underscore the immediate use of assessment information to feed back into the instruction and learning process is essential. Please let me put forth AASA's proposal to improve both the information used by teachers and information desired by policymakers.

Rather than a national test, we need a consensus on what students are expected to know and be able to do. The National Council of Teachers of Mathematics have developed a splendid set of standards for mathematics that serve as an example, and I can say that in our school district we have just this year implemented a curriculum, K-12, using those standards. The science, English, social studies and reading teachers have and are doing similar work.

Congress should capitalize on these efforts as well as the work of the National Assessment of Educational Progress and various academic commissions to develop a national consensus on standards for education. And by standards, I am talking about expectations, not something that is a rigid level for all school districts. Because I don't think, personally, that you can come up with something that will fit every school district in this Nation.

We do not need a single list of desired outcomes, but rather the national standards framework called for by Lauren Resnick recently in testimony to the Romer panel on implementation of the national goals. These standards are most definitely not a curriculum, and most emphatically not a test. Rather, they describe what students should know and be able to do. States and school districts can then determine how to achieve the standards and which standards should be emphasized in that particular setting.

If employers could have accurate and easily understood descriptions of what students knew and could do, and if all students had achieved high standards, there would be little use for other test information. I have personally told business people in our community that they do not use our quality control mechanism which has existed for a long time; namely, grades.

It is extremely unusual—and I suspect if you go back to your communities you will find the same thing if you ask the principals. It is extremely unusual for an employer to ask for a student's transcript when he or she is being hired. This would tell the employer a great deal about the degree of success of the individual and reinforce for students that their efforts in school do count.

Now the reality is that when a businessman says to me, "I've got a young woman from your high school, Karl, and she can't type," I talk to them about this topic, and I say, "Do you know whether she was a D student in typing, or a C student, or a B student, or an A student?" If she was an A or B student and she can't type, that is news to me. If she got a D-minus, we all knew; we knew it and you could know that she wasn't one of our good ones, you see. And so the reality is we don't use the mechanism that we have for quality control of individuals as it is.

In the past, I have spent a great deal of time listening, thinking and talking about this, particularly in the last year, especially the subject of assessment of children's progress. It is my feeling that there is much to be gained in the instructional arena from assessing the progress of children in a way which can be, again, underlined for you, immediately fed back into the improvement of instruction.

However, I must say that I personally feel that society has pinned entirely too much of its hopes for the improvement of American education on the testing of children. In some way I sense that we hope that by having great national testing of children that we will embarrass ourselves into being better.

Then there is the unanswered question in my mind of whether there is linkage on the national scene between these tests and other agendas such as choice. There are other facets of the education of children which it seems to me are far more at the core of the improving of the learning of children such as (1) improving the health, family situation and learning of preschool children in disadvantaged situations; (2) demanding an acceptance by all of the educators that every child is capable of learning no matter what the youngster's background is and that we do not automatically categorize some students as being academically deficient because of their background, and that is happening; (3) convincing ourselves as educators that all of these children can be lifted to a high set of expectations and that we will demand the very best of them; and (4) revolutionizing parents' expectations that would lead them to accept as Asian parents do that hard work of children is the primary factor in their success in schools. That is not what American parents are telling us. They are not telling us that hard work is at the core of their children being successful. They are telling us that it is what God happened to give their kids; in other words, they have either got it or they haven't got it, which in my mind takes everybody and the mechanism off the hook: the child is off the hook, the teacher is off the hook, the parents are off the hook; whereas, if you have the mind-set that hard work is going to do it, then we are all on the hook. And (5) making a national commitment that we would extend ourselves mentally and financially to improve education.

On behalf of AASA, I thank you for the opportunity to express our views on national testing, assessment, and accountability.

[The prepared statement of Dr. Karl V. Hertz follows:]

AMERICAN ASSOCIATION
OF SCHOOL ADMINISTRATORS

TESTIMONY

OF

DR. KARL V. HERTZ
SUPERINTENDENT
MEQUON-THIENSVILLE SCHOOL DISTRICT
MEQUON, WISCONSIN

ON

NATIONAL TESTING/ASSESSMENT AND ACCOUNTABILITY

TO THE

SUBCOMMITTEE ON ELEMENTARY, SECONDARY,
AND VOCATIONAL EDUCATION

UNITED STATES HOUSE OF REPRESENTATIVES

THE HONORABLE DALE E. KILDEE
CHAIRMAN

THURSDAY, MARCH 14, 1991

WASHINGTON, DC

(703) 528-0700 • Fax (703) 841-1543
1801 North Moore Street • Arlington, Virginia 22209
An Equal Opportunity Employer

Mr. Chairman and members of the subcommittee, on behalf of the American Association of School Administrators (AASA), I would like to thank you for the opportunity to testify on this most important issue. AASA is the professional organization of over 19,000 school district superintendents, other local administrators, and professors of educational administration. In our capacity as education CEO's in our communities, we have a vital interest in the discussions regarding possible development of a national test.

I am superintendent of the Mequon-Thiensville School District in Mequon, Wisconsin which is a suburb of Milwaukee. Our district has an enrollment of 3,500 students employing 240 professional staff. During the past year, I have served on the Wisconsin Governor's Commission on Schools for the 21st Century which flowed from the president's gathering of governors regarding education. I chaired the work on assessment.

AASA would like to address three of the issues surrounding the development of a national test.

CAN A NATIONAL TEST BE THE MAJOR FORCE IN IMPROVING EDUCATION?

The short answer is no, a national test cannot be the major force in producing educational quality. First, a national test would hold students responsible for conditions they do not create or control. In general, children and youth who live in poverty score worse on tests than economically advantaged students. Children whose lives are wracked by poverty, family tragedies, and major personal problems often do not do well on tests unless the underlying causes are ameliorated. Similarly, we cannot say with certainty what effect pervasive racism and accompanying low expectations have on student performance, but it must be considerable. An associated dramatic concern connected to national testing is the inappropriate placement of disadvantaged children, especially those who come from non-standard English speaking homes, in classes that do not challenge their true talents.

Second, a national test would occur separately from the daily learning process. Tests have been asked to fill two needs, instructional information and public accountability. No current test is able to fill both needs completely although some states are working toward that end. Because the test would be removed from the instructional setting, the results are unlikely to be important to students. Experience in state testing shows that tests which are viewed as unimportant do not elicit the best efforts of students. Another low interest factor for students is related to the fact that they are separated in time from the test, further reducing the test's significance to students. As Lew Rhodes of the AASA staff recently wrote, "By removing assessment from any functional role in the instructional process, we have virtually made it impossible for teaching to be a continuously improving process under control of a teacher; or more importantly, for learning to be a continuously improving process under control of the learner." Further, the Wisconsin Commission's report identified the relationship between the assessment process and the immediate and direct application of that information back into the instructional process. To teach children better now is the number one purpose of assessment. If testing is not directly and immediately linked to improved instruction for children, it has little chance of having impact.

Third, the notion that quality can be achieved chiefly through an "end of the process inspection" is wrong. Quality cannot be inspected into a product; rather, quality can only be built in. Perhaps the committee should invite the father of total quality management, W. Edwards Deming, to explain that principle and forever put to rest the myth that any type of end of process inspection, including tests, will result in quality for any activity, including education.

Fourth, developing a single national test will take a long time, and given the need to hold down expense, the test will likely have to be a multiple choice or short answer test. Test makers are working on improved norm referenced and criterion referenced standardized tests, but tests now rely on multiple choice or short answer test items rather than demonstrations of learning. Waiting on a national test to solve educational problems, and then discovering that testing will not improve the system will take years, maybe decades. We simply cannot take that much time. I would also suggest that the construction of a national test will be fraught with political struggles which will neutralize any potential value it might hope to have.

Fifth, a multiple choice short answer test will almost certainly fail to accurately measure student learning because a national test will not test what is being taught. Also, work on learning styles and multiple intelligences shows clearly that how questions are asked has a great deal to do with how well students perform. No single test can respond to the learning styles found in any large group of students. Finally, since there is no national consensus about what students need to know and be able to do, no national test has any basis for choosing what to test.

CAN TEST SCORES FORCE EDUCATORS TO DO A BETTER JOB?

Again, the short answer is no. Teachers and administrators whose students do not score well can be humiliated, but in the absence of clear information about the causes of failure to learn, improvements cannot be made. The notion that humiliation can produce improved results assumes that teachers and administrators are holding something back in their professional practice that they know how to do better but for some reason will not do. Every study of teachers and principals has found that most educators do their best each day. An assumption that educators know something they are not using has no supporting evidence.

Teachers and administrators must improve their competence in developing assessements and using assessment results to facilitate student learning and improve their teaching. A national test developed by either the federal government or some private group would not add to this pressing need.

IS THERE AN ALTERNATIVE TO A NATIONAL TEST THAT CAN PROVIDE INFORMATION ABOUT WHAT STUDENTS KNOW AND CAN AT THE SAME TIME BE USEFUL TO BOTH THE INSTRUCTIONAL PROCESS AND POLITICAL LEADERS?

The short answer is an emphatic yes. A number of states are now developing information systems about student learning that can serve both the need for daily information on what has been learned and the need of political leaders to understand the impact of educational policies. Vermont, California, Connecticut, and Arizona come to mind as states working on learning information systems that can serve all masters. Even more hopefully, several hundred school districts have been developing powerful information about student learning for use by students, teachers, parents, and administrators.

School districts that have made great strides using information describing student learning and mastery have several traits in common. First, such school districts began by clearly describing what students are expected to know and be able to do when they finish school. Second, the districts did a careful study of how they could best facilitate the desired outcome through structure and curriculum. Third, they spent considerable time in teams deciding what evidence of progress and mastery would best facilitate learning and improve teaching. Finally, the school districts carefully studied what more they needed to know and be able to do to teach every student the desired outcomes and use information about learning to continuously improve.

School districts engaged in fundamental change have learned that change takes time and that they must strive for constant improvement. There are no quick answers in education or any complex human activity. The challenge is to accept the principle of continuous improvement rather than try massive instant change.

The transformation depends upon information about student learning and behavior that first and foremost facilitates student progress and the professional practices of educators. However, aggregate information on students' gains can be described in terms of the locally developed outcomes, state-wide tests, or commercially available tests.

There are many types of information about student progress, some of which can be aggregated upward and used throughout a system in the sort of examination system envisioned by Lauren Resnick. But first, there must be a consensus on what students are expected to know and be able to do, otherwise there is no basis for judging growth. Only after expectations have been defined is it reasonable to ask what evidence is acceptable that learning has taken place.

The fact is that different districts in different locations and at different points in history have varying expectations, goals, and aspirations. When a national test is given in such a setting, it can quickly become a common level of minimum competency which can cause huge numbers of people to focus upon an expectation which is considerably out of sync with the condition in a given community.

During the deliberations of the Wisconsin Commission on Schools for the 21st Century, the Assessment Committee heard again and again that norm referenced tests were not used to improve the instruction of children. It became clear to us that judging the success of children could be done in a variety of ways. Clearly, by looking at the writing, portfolios, experiments, and projects of children, teachers are particularly able to take what they have learned about the students' progress and feed it directly back into the instructional process within a very short period of time which is essential to making the learning effective. This can actually be done with norm referenced tests too. However, the timing is once again delayed to a certain extent, and, of course, in any kind of national test it would be delayed dramatically and separated from the learning process to an even greater degree. When our district receives the results of national testing, or is doing assessment of writing or any of the other more immediate forms of assessment, we take the results into small groups of teachers and building administrators and plan exactly what the educational strategy will be to effectively use that information to better instruct the children. In some cases, this may result in changes which will take place immediately, in other instances it may mean that the topic will be approached differently in another year, and in some cases the team of educators will decide that the children did poorly but we did not intend to teach that topic, and so we try to avoid a precipitous reaction to try to address a situation that we may know is covered very appropriately at some other point in the child's learning situation. We feel that an immediate use of the assessment information to feed back into the instruction and learning process is essential.

Please let me put forth AASA's proposal to improve both the information used by teachers and information desired by policy makers. Rather than a national test, we need a consensus on what students are expected to know and be able to do. We need a massive commitment to improved practices regarding student assessment so that the question of acceptable evidence can be answered to the satisfaction of business people, parents, educators, and the community at large.

The National Council of Teachers of Mathematics has developed a splendid set of standards for mathematics that serve as an excellent example. The science, English, social studies, and reading teachers have or are doing similar work. Congress should capitalize on these efforts as well as the work of the National Assessment of Educational Progress and various academic commissions to develop a national consensus on standards for education. Because all of the work on standards has been done in small elite professional groups, a national consensus on the important next step has eluded us.

Congress should act quickly to establish a broadly based coalition to achieve the necessary consensus on what we expect students to know and be able to do. This should be done in conjunction with the Romer panel, but without repeating the mistake of only talking with a narrow band of individuals. Rather, the congressional panel should spend one or two years achieving the consensus on standards by building on the work already done, being inclusive, and by actually listening and achieving consensus on what parents, business, higher education, and the public at large actually want.

We do not need a single list of desired outcomes, but rather the "national standards framework" called for by Lauren Resnick recently in testimony to the Romer panel on implementation of the national goals. These standards are most definitely not a curriculum and most emphatically not a test. Rather, they described what students should know and be able to do. States and school districts can then determine how to achieve the standards and which standards should be emphasized. After there is a consensus on expectations, the questions are, first, how do we achieve those expectations as soon as possible? Then, what evidence of student progress and mastery are required to facilitate continuous learning and improve educational practice and structure? Finally, the question of how to assess the evidence of performance success should be answered.

If employers could have accurate and easily understood descriptions of what students knew and could do and if all students had achieved high standards, there would be little use for other test information. I have personally told business people that they do not use our quality control mechanism which exists - namely grades. It is extremely unusual for an employer to ask for a student's transcript when he or she is being hired. This would tell the employer a great deal about the degree of success of the individual and reinforce for students that their efforts in school do count.

Next, Congress must appropriate funds to support the development of competence in student assessments and in using that information with students. Students use information in all other aspects of their lives to make decisions; they can use assessment information just as well.

In the past year, I have spent a great deal of time listening, thinking, and talking regarding the subject of assessment of children's progress. It is my feeling that there is much to be gained in the instructional arena from assessing the progress of children in a way which can be immediately fed back into the improvement of instruction. However, I must say that I personally feel that society has pinned entirely too much of its hopes for the improvement of American education on the testing of children. In some way I sense that we hope that by having great national testing of children that we will embarrass ourselves into being better. Then there is the unanswered question of whether there is linkage on the national scene between these tests and other agenda such as choice.

There are other facets of the education of children which it seems to me are far more at the core of improving the learning of children such as: one, improving the health, family situation, and learning of pre-school children in disadvantaged situations; two, demanding an acceptance by all of our educators that every child is capable of learning no matter what the youngster's background is and that we do not automatically categorize some students as being academically deficient because of their background; three, convincing ourselves as educators that all of these children can be lifted to a high set of expectations and that we will demand the very best from them; four, revolutionizing parents' expectations that would lead them to accept, as Asian parents do, that the hard work of children is the primary factor in their success in school; and, five, making a national commitment that we would extend ourselves mentally and financially to improve education.

On behalf of AASA, I thank you for the opportunity to express our views on national testing, assessment, and accountability.

Chairman KILDEE. Thank you, Dr. Hertz.

Dr. Neill?

Dr. NEILL. Thank you, Mr. Chairman, members of the panel. Thank you very much for inviting FairTest to present testimony here today. My name is Monty Neill. I am the Associate Director of FairTest, which is the Nation's only organization whose sole purpose is to look at and try to ensure that the assessment of our Nation's students and workers is made fair, open, relevant, and indeed, helpful. I would like to summarize my written testimony.

Based on our examination of existing proposals, FairTest urges the House to support education reform by not implementing a national examination or examination system at this time. The House should, however, support efforts to introduce new assessment methods as part of improving school quality.

National testing proposals assume that more measurement will produce positive change. Recent history shows this is not true. During the 1980s U.S. school children became the most overtested students in the world, but the desired improvements did not occur. Adding more testing is clearly not the way to improve education anymore than taking the temperature of a patient more often will reduce his or her fever.

Successful educational reform must begin by defining the kind of education we want our children to have. Not simply standards, but what experiences do we want our children to live through every day when they go to school. On that basis, we can then determine the changes in standards, curriculum, school governance, and assessment required to reach the educational goals, and on that basis we can then decide whether we want or need a national assessment system.

Current proposals for a national test or examination system seek to test before the necessary decisions about the goals and methods of school reform have been made. This is putting the cart before the horse. Some current proposals, such as the one presented by Mr. Cooperman from Educate America, call for the creation of a test that would cost about $30 for six tests, to be administered to all students sometime in the near future. Because of cost and time factors, such a test inevitably will be entirely or mostly multiple choice. To do a performance-based test, let me give you an example. The ART Advanced Placement Test cost between $60 and $70 a year. If you multiply that times 6 times 3 million students, we are spending on the order of a billion dollars a year to have a test comparable in quality to the ART AP Test, not any $30 per student.

In my written testimony I have explained why the imposition of a national multiple-choice or short-answer-type test will be an educational disaster for the country. It is because of their limitations as technical instruments and because of their very harmful effects on curriculum and instruction. Yesterday, Representative Goodling read into the record a letter from the Pennsylvania Secretary of Education that I wish I had written myself, because it summarized extremely well the drawbacks of that kind of testing.

No proposal that relies predominantly or very much on multiple choice testing should be given any serious consideration by the House. Because the National Assessment of Educational Progress, about which you heard yesterday, remains predominantly multiple choice, any effort to transform it into a national test will have similar disastrous effects on educational reform. NAEP should remain what is, a national indicator, but should move toward more performance-based assessment methods.

Now what about implementing a national performance-based examination system, which is really the primary discussion that went on in the previous panel. It certainly has the advantage over multiple-choice testing in that it uses methods that can assess high order thinking abilities and that support good educational practices that in turn could develop the creativity, problem-solving abilities, reasoning, cooperative work habits, and so forth, of the students that I think we want to attain.

However, it is our belief that we can move toward the national use of such assessments without constructing a national examination system. We would then gain the advantages of good assessment and avoid the disadvantages of imposing a national examination system.

Testing by itself of any type, getting information by itself of any type in itself will not produce reform. Information has to be used in the context, a much broader of—that is, outcome information from students has to be put in a much broader context of knowing about the inputs, the social factors around the school, the processes that go on in the school, and indeed other output measures. So, if we focus only on testing and assessment as a solution to our educational problems, we will end up failing to address such critical issues as equity, rigid and bureaucratic school governance and structure, low quality textbooks, inadequate schools of education, a lack of information about everything else we need to know about.

We would be also imposing on teachers a new examination and the teachers will not be able at this time to use them. To make real use of performance-based assessments requires creating performance-oriented schools, which in turn requires restructuring, staff development, and new educational materials. In short, we would be ill-advised to jump first to a national test before figuring out the other things that we need to do. We must organize the pieces of a program of school reform into a coherent whole, rather than jump at one part.

We also don't know at this time whether it is feasible to construct a national examination system of the sort described earlier. The whole progress could prove too complex, expensive, and unwieldy to work. Britain, for example, my understanding is, has just dropped moderation from its national system. Now, moderation is a process whereby the teachers grade the examinations but they do so in a process of constant discussion among each other in order to constantly redefine what is good work as distinguished from average work, as distinguished from unacceptable work. It is also a process that feeds back into redesigning the assessments. It is a process whereby teachers learn a lot. It is also a process by which the curriculum can be restructured. It is, in short, crucial to a performance-based examination system. But, as I said, Britain dropped it, largely because it is too expensive. Britain does not test remotely the number of students that we were talking about testing.

FairTest also is very concerned about issues of bias and fallibility. On any examination, there will be students who fail who should have passed, and experience tells us, from the history of examinations in this country, that a student from low-income and minority group backgrounds will be disproportionately among the students who fail the tests but should have passed the tests, and who therefore suffer unfair negative consequences from testing.

While I am sure that the proponents of a national examination system do not support tracking or grade retention, we have no particular reason to believe that by setting up a situation whereby students are expected to attain a certain level of mastery by the age of 16 that it will not in turn feed back to a process of tracking and frequent retention. Indeed, the discussion earlier suggested that students who don't attain a certain level at grade 4 might be retained. We know about as well as we know anything about education that retention in grade simply does not work.

I like the idea of more time in order to accomplish what is needed, but it can't be done by retention. The message then is we have to think about restructuring things like grade levels and things like use of time in the school day. We have experiments around the country on how to do that. We need to learn a lot before we drive everything with a national examination.

Instead of implementing a national exam at this time or a set date such as the year 2000 by which we can expect such an examination system to be in place, FairTest urges the Federal Government to take the following steps to improve educational assessment.

First, assist States and districts acting in consortia to develop and implement performance-based methods of assessment. For Example, the various consortia working on new assessments such as the Learning Resource and Development Center at Pittsburgh and the National Center for Education and the Economy, as well as the consortia from the Chief State School Officers, another one from the Education Commission of the States, another one out of Boston College headed by George Madaus from the National Commission on Testing and Public Policy—they have efforts under way. Their efforts should be supported. This means funding research and pilot program implementation.

You should also assist State and district teacher education and staff development programs because if teachers are not able to use the new assessments, the new assessments will not help us improve education.

You could assist the subject area groups to develop and disseminate model frameworks, curricula, standards, and assessments.

Also, assist all of these different levels of groups in sharing information among themselves, such as the NCTM standards, how do they get widely distributed around the country and how do teachers and administrators learn to use them?

The standards we are talking about, I would add, are ultimately what the Coalition for Central Schools refer to as habits of mind, and what a group I am now working with has begun to call intellectual competencies. It is not simply bits of information that people need to know, nor are they fixed things you have to be able to do at a certain time. Rather, it is, for example, our history students should learn to think the way historians think. Now there is a lot of levels at which that can be done and we should, indeed, set levels that our students are shooting for and doesn't prohibit much higher levels such as that attained by professional historians. But that is the approach we need to take that is a radically different approach than most of our schools in this country use and it cannot be measured with most of the current testing instruments that we have.

We also have to be sure that our assessment plans are embedded in thorough educational reform, and we urge you not to act unless you know they are. Don't just assume that they will be.

I would also urge you to reconsider how the Federal Government now mandates standardized testing such as with Chapter 1. All across the country I hear repeatedly what a disaster those mandates are for good education in Chapter 1 programs.

In conclusion, let us say we are not arguing at FairTest against accountability or for slowing down school reform. Rather, the central issue is how do we define education? How does our Nation define education? We need school reform, not more testing. Our Nation must not be misled into thinking that more testing will solve our educational problems. Indeed, good assessment is part of the solution, it is not the first step toward a solution. We will be far better served to take the time to do the job of school reform carefully and well than to act hastily and poorly with destructive results.

Thank you very much.

[The prepared statement of Dr. Monty Neill follows:]

FairTest
National Center for Fair & Open Testing

Why the United States Does Not Need a National Test:

Testimony to the House Committee on Elementary and Secondary Education

by

Monty Neill, Ed.D., Associate Director,

National Center for Fair & Open Testing (FairTest), Cambridge, Mass.

March 14, 1991

Ladies and Gentlemen,

Thank you very much for inviting FairTest to appear at this important hearing.

Based on an examination of existing proposals, FairTest concludes that most current efforts to establish a national test to measure progress toward the nation's educational goals will hurt, not help, our nation's efforts to improve school quality. We therefore urge the House of Representatives to support education reform by ***not*** implementing a national exam at this time. The House should, however, support efforts to introduce new assessment methods as part of implementing school reform.

National testing proposals largely are based on the false premise that measurement by itself will produce positive change. Recent history shows this is not true: During the 1980s, U.S. school children became probably the most over-tested students in the world -- but the desired educational improvements did not occur. FairTest research indicates that our schools now give more than 200 million standardized exams each year and the typical student must take several dozen before graduating.[1] Adding more testing will no more improve education than taking the temperature of a patient more often will reduce his or her fever.

In contrast, successful educational reform must include restructuring curriculum, instruction, textbooks and other materials, school governance, and teacher education, as well as assessment. What we need to create are schools as communities of and for learning.

To move toward that goal, teachers, administrators, other school personnel, parents, students, community members, and government must all be involved in an open and democratic process of defining our educational goals -- at the local, state and national levels -- so that we can agree, for example, on what it means for all students to be competent in different areas. On that basis, we can then determine how to make the changes required to reach the goals, including a decision on whether to institute a national test. Most current proposals for a national test, however, seek to test *before* necessary decisions about the goals of school reform have been made. This likely will lead to the backdoor imposition of a national curriculum, without public discussion.

Indeed, having a single national test raises the issue of the control of education. If the test becomes important, as all testing proponents want, those who control the test could control curriculum and instruction, particularly if decisions about curriculum and instruction have not been arrived at before the test is constructed, and maybe even if those decisions have been reached.

A national exam should not be allowed to undermine such needed and emerging reforms as school-based management and shared decision-making. By centralizing decision-making, centralized national testing most likely will make education less, not more, accountable to parents, students, teachers and the community. If the test is centrally controlled, to whom could parents, teachers and communities appeal if they disapprove of the curricular decisions and instructional methods imposed through the test?

Any national test or examination system could produce additional major problems. For example, the examination systems of many other nations are based in school systems that rigidly sort their students and perpetuate social class divisions, precluding the social mobility our nation believes in. No one test should become a national gatekeeper that perpetuates our nation's unfortunate history of unfairly sorting students by race and class.

A national test could end up being used to determine high school graduation, employment and entrance into higher education. Due to unavoidable measurement error and bias, many students who fail a test will, in reality, be as capable as many who pass. Research indicates that those who fail but should have passed will be disproportionately from low income and minority group backgrounds. FairTest agrees with the National Commission on Testing and Public Policy that, because of the bias and error, no one test should ever be the sole or primary basis for making an important educational decision.[2]

Dangers of Multiple Choice Testing

FairTest recognizes that there are two different types of proposals for national testing. One type will rely essentially on multiple-choice testing; the other calls for performance-based assessment. These two approaches are quite different. They are the difference between testing *what students should know* and *what students know how to do*.[3]

The first approach quickly leads to multiple-choice testing of arbitrary facts and isolated skills, unconnected to the way knowledge is used in the world. Multiple-choice and short answer tests cannot adequately assess problem-solving or the ability to create and use knowledge.[4] Higher order thinking requires the student to define the problem, to consider and attempt various solutions to problems which are ill-structured and may have more than one correct solution, and to produce knowledge, not merely recognize answers.

Because multiple-choice/short answer testing cannot directly assess higher order capabilities, a test comprised of such items will not inform us as to the problem-solving and knowledge-creating capabilities of our students. We know from research, however, that student abilities in these areas are very limited. This has been caused largely because schools of schools' failure to teach them in any subject area to more than a few students. Even the best high school students typically do not know how to problem-solve using the approaches and methods a professional uses.[5] Yet research also shows that problem-solving, knowledge-creating approaches can be used even with very young children.[6]

If a test is important -- as a national test is sure to be -- then teachers will teach to it.[7] Because multiple-choice tests cannot directly measure higher-order skills, teaching to the test reduces or eliminates instructional time spent on the higher skills. Instruction is reduced to drilling for multiple choice exams and the curriculum is reduced to the test. Multiple-choice testing precludes a curriculum based on thinking, investigating, problem-solving and using creativity, because the test cannot measure those things.

Additionally, these tests perpetuate the false idea that first students learn basic skills, then they learn higher skills. Cognitive psychological research has demonstrated that learning involves active thinking and to enhance learning the student must be actively engaged.[8] Test-driven schools produce higher test scores, but not students who are able to think.

A predominantly multiple-choice test may include a writing sample. A typical short writing sample requires a student to write several hundred words on a topic he or she may or may not know anything about and may or may not care about, in a short period of time, with no chance for research, discussion (that is called cheating), or serious revision, for no purpose except the test. If the purpose of writing is to communicate, then a typical test writing sample cannot legitimately be called writing at all. As with multiple-choice testing, it sends the wrong message about the goals of education.

These tests are not very useful to teachers or policymakers. The reason, in both cases, is that the test results do not help the teacher or policymaker decide what to do next. If Johnny cannot multiply, the test cannot explain why. If Maria's whole class cannot multiply, the test does not provide information on what should be done.

What standardized multiple-choice tests do best is help sort students: the good, the bad and the ugly. It is what they were invented to do. But if we are serious about reforming education so that all students can learn the things we deem important, then we must stop relying on tests that have as their only real use the sorting of students.

In sum, implementing a national multiple-choice exam will mislead the public about the nature of the problem and the requirements of real change, block positive school reform (including the use of new methods of assessment), hinder students' ability to develop the kinds of intellectual competencies they need to develop, and ultimately undermine public education.

No proposal that suggests using more than a small proportion of multiple-choice items in a national examination should be given any serious consideration by the House. At most, multiple-choice could be used as part of a sampling program to gather limited information about student acquisition of a narrow range of knowledge. There is no reason to test every student for this purpose and such a purpose should never be allowed to dominate education, as it too-often now does.

Because of current technical limitations, any proposal to assess our nation's students inexpensively and in the near future will, of necessity, be a multiple choice test. An example is the proposal by Educate America to test all high school seniors in six subjects for $30 - $50 each. Such proposals must be rejected.

Performance-Based Assessment

By contrast, students should be assessed on what they know how to do. To know how to do something includes knowing factual content. This method of assessment corresponds to how people learn. They learn by integrating new information or experiences into the intellectual frameworks they already possess, which in turn enables them to refine and improve the frameworks.

Assessing what students know how to do is based on students' doing real work. There are many ways for students to demonstrate intellectual competence in and across the subject areas. Performance-based assessments can be based on regular student classroom work - projects, research, writings, products, self-reflection, teacher evaluation, exhibitions, and performances - that can be organized and summarized in portfolios. In turn, the portfolios can be examined by outside people -- teachers, other parents, trained examiners -- to determine the quality of the portfolios and the kinds of work students are doing. Vermont for example, is working on this method.

Performance-based assessments can also be examinations administered from outside the classroom. These can include open-ended, complex problems requiring the student to figure out what to do, solve the problems, and explain what he or she did. Or they can be exhibitions, performances and products, such as now done in science fairs, Scout Merit Badges, Advanced Placement Art, and many performing and applied arts. These often can be exams that are worth teaching to, unlike multiple-choice tests. Arizona, California, Connecticut and Maryland are among the states implementing these types of exams.

Taken together, in-class and externally-developed performance-based exams can encourage real work, model high standards, spur improvements in teaching and curriculum, produce instructionally useful information for teachers and students, and provide information based on real activities about student progress. Assessment can play an important part in developing communities of and for learning.

Cautions on a National Performance-Based Examination System

However, support for performance-based assessments does not mean such assessment should immediately be transformed into a national examination system, such as that proposed by the Learning Research Development Center and the National Center for Education and the Economy (LRDC/NCEE).[9] There are many reasons why this is the case. Among them are:

-- We have not yet completed the process of discussing and debating what we want our educational systems to be. Many complex issues of educational reform, involving curricular goals and standards, instructional methods, assessment methods, school structure and governance, and collection of information, largely must be resolved *before* the question of whether a national examination system is desirable can be answered. To do otherwise is to put the cart before the horse.

-- Imposing a national examination will not address the issues of rigid and bureaucratic school governance and structure, low-quality textbooks, and inadequate schools of education. Improving assessment needs to be considered as one part of integrated systemic change.

-- The proposal calls for national boards to set standards. It could create a national school board that, by setting curriculum standards, will lead to a centralized, national education system. Because the consequences of such actions cannot now be known, but may include undermining democratic control of education, we should not rush into that process.

-- Staff development is central to school reform, but is not adequately addressed in the LRDC/NCEE proposal. If teachers are to teach to performance-based assessments, to teach the "thinking curriculum," they have to know how to do so. This involves developing the ability of our nation's 2-1/2 million teachers to teach and assess in new ways. To be effective, school reform must include the active participation of those who will implement the changes. We cannot impose new assessments on teachers, change nothing else, and say "Do it."

-- In general, the proposal does not adequately address equity issues that must be solved for the system to be fair. Changing assessment will not by itself reduce inequities. All students must be assured a fair opportunity to learn how to work within a thinking curriculum that uses performance-based assessments. Additionally, the goal of "initial mastery" could encourage sorting and tracking students according to who can best or most quickly reach the goal. This danger needs to be seriously addressed to try to ensure structures and processes, including in the realm of assessment, that are inclusive and reduce tracking and other kinds of sorting.

-- We simply do not know whether it is feasible to construct a national examination system. The whole process, particularly the calibration, could prove to be too complex, expensive and unwieldy to work. For example, England recently dropped a moderation process from its national exam process because it was too expensive. Moderation is the process by which teachers help shape standards and learn to grade papers, products and performances uniformly so as to produce consistent and reliable results. Moderation is valuable and necessary and must be included in any performance-based system, but doing it on a national level on top of state and local levels may be too much as well as unnecessary for educational improvement.

-- When the complexities and expense of the proposal become clear, the portfolios and projects could end up being reduced to very limited exams. There even could be a return to multiple-choice and short-answer exams. Such a retreat would have destructive curricular effects and undermine all aspects of educational improvement.

-- The proposal is not conceived of as one part of an overall educational information system. Having assessment outcome information on education is not useful unless we also have adequate information on inputs (money, teaching staff, building quality, etc.), processes and programs (curriculum, instructional methods, textbooks and materials, class size, role of tracking, governance and school organizational structure, etc.), and additional outcome data (employment and further education of graduates, dropout rates, etc.). This information should be obtained without harming education -- unlike what has happened with multiple-choice tests. Schools and programs should be evaluated on a comprehensive range of indicators of their quality as communities that support learning for all students.

Recommendations

There is no one, simple method of putting a national education reform process into motion in the right direction. It is a process that can and is happening at all levels: the classroom, the school, the district, the state, consortia that include all of these, and at the national level. It is not and will not be a smooth and easy process. But as good practice becomes available to replicate, as improved curriculum and assessments become more widely known, as our nation's desire to improve education for all continues to grow, then we can expect to see real progress.

The federal government can proceed in one of two ways. It can impose a national test that runs the risk of short-circuiting the process of school reform. Or it can find ways to support school reform activities without imposing a national test.

FairTest concludes that the House of Representatives should not propose a national exam either immediately or to be in place within any fixed timetable, such as five or ten years. Rather, FairTest urges the federal government to take the following steps to improve education and assessment:

-- Assist states and districts, acting in consortia, to develop and implement performance-based methods of assessment.

-- Assist state and district teacher education and staff development programs.

-- Assist the subject area groups, such as those in math, English, social studies and science, to develop and disseminate model curricula, standards and assessments.

-- Re-examine the instances in which the federal government requires standardized multiple-choice testing, particularly for the Chapter I program. The testing requirements virtually force programs into being test-coaching programs, though that, as explained above, is a poor educational method.

-- Leave the National Assessment of Educational Progress as a national indicator.[10] To turn it into some kind of a national test will end up destroying its current usefulness and may produce all the drawbacks discussed above. In particular, NAEP should not be used below the level of state-level comparisons. FairTest doubts that state-level comparisons will be of real use to educators and urges that state comparisons not be approved beyond trial measures unless experience and research demonstrate how the comparisons will be used to improve education. NAEP should, however, include far more performance-based assessments and provide technical assistance to districts, states and consortia who are implementing performance-based assessment.

-- Consider how assessment information can best be included as one element of school reform activities and one part of an indicator system, and not view assessment in isolation.

Only after these educational reform processes have been implemented and evaluated over a period of time should the federal government consider whether it is desirable or feasible to link the newly developed local and state performance-based assessments to each other and to national standards or curricular frameworks.

Let us be clear. FairTest is are not arguing against accountability or for slowing down school reform. Nor is the issue one of the need for "standards." Rather, the central issue is how we define education. We are saying that we need school reform, not more testing. We need genuine accountability, not test scores from multiple-choice or short-answer exams, and we don't need to jump aboard an examination train heading into trackless terrain.

Our nation must not be misled into thinking more testing will solve our educational problems. Instead, we must construct plans for reform that include assessments which can be used to help student learning, guide educational improvement, provide information for accountability, and assist the goal of equity, but not block progress or harm students. Our nation will be far better served to take the time to do the job well, than to act hastily and poorly with destructive results.

Endnotes

1. Medina, Noe and D. Monty Neill, *Fallout from the Testing Explosion: How 100 Million Standardized Exams Undermine Equity and Excellence in America's Public Schools* (Cambridge, MA: FairTest, 3rd Ed., 1990).

2. National Commission on Testing and Public Policy, *From Gatekeeper to Gateway: Transforming Testing in America* (Chestnut Hill, MA: Author, 1990).

3. Susan Harman, drawing particularly on the work of people associated with the Coalition for Essential Education, uses this formulation very clearly in "National Tests, National Standards, National Curriculum," *Language Arts* (January 1991: 49-50).

4. For a general critique of multiple choice testing, see Neill, D. Monty and Noe J. Medina, "Standardized Testing: Harmful to Educational Health," *Phi Delta Kappan* (May 1989: pp. 688-697), and Medina and Neill, *op. cit.* For analysis of why multiple-choice testing cannot assess higher order thinking, c.f. Fredericksen, Norman, "The Real Test Bias: Influences of Testing on Teaching and Learning," *American Psychologist* (March 1984: pp. 193-202); National Commission on Testing and Public Policy, *op. cit.*; Resnick, Lauren B. and Daniel Resnick, "Assessing the Thinking Curriculum: New Tools for Educational Reform," in B.R. Gifford and M.C. O'Connor, eds., *Future Assessments: Changing Views of Aptitude, Achievement, and Instruction* (Boston: Kluwer Academic, 1989).

5. "Beyond the Bubble," a mini-conference at the April 1990 national conference of the American Educational Research Association, Boston, MA.

6. Kamii, Constance and Mieko Kamii, "Why Achievement Testing Should Stop," and Engel, Brenda S., "An Approach to Assessment in Early Literacy," in Kamii, Constance, ed., *Achievement Testing in the Early Grades: The Games Grown-Ups Play* (Washington, DC: National Association for the Education of Young Children, 1990).

7. Madaus, George. "The Influence of Testing on the Curriculum." *87th Yearbook of the National Society for the Study of Education. Part I: Critical Issues in the Curriculum* (1988: pp. 83-121). Also, Resnick and Resnick, *op. cit.*

8. Resnick and Resnick, *op cit.*

9. Learning Research and Development Center and National Center on Education and the Economy. *Setting a New Standard: Toward an Examination System for the United States. A Proposal* (Pittsburgh and Rochester: Author, October 1990) is the most comprehensive of the reports and documents proposing a national performance-based examination system.

10. Campaign for Genuine Accountability in Education, "Open Letter to Congress, Bush Administration, the Governors on NAGB and NAEP Expansion," (Cambridge, MA: FairTest, June 15, 1990).

Chairman KILDEE. Thank you very much.

Our next witness is Dr. Kelley.

Dr. KELLEY. Mr. Chairman and members of the subcommittee, I thank you for having me. I am going to approach them from, hopefully, two different perspectives other than a national one. I am on a school board right now, and its past president. I have been in, I think it is fair to say, hundreds of school districts in six countries on three continents, and hundreds of classrooms as well. I can give you my perspective better by telling you what I am going to do the next 10 days.

When I leave here I am going to Schenectady, New York, and work with their board and their administrators to reorganize their district in two senses: they have to redistrict, and they are consolidating a high school. With a little luck and help, we will be able to induce and encourage teachers in that high school to take accountability for that product. That is very unusual in a high school in the United States.

After that I will have to go England. It turns out that the English have decided this spring to test all 7-year-olds. I am not sure what they are going to test them on, but politicians being much more practical than, perhaps, people involved in testing, all of a sudden raised a question. The minister asked, "All right. We are going to test these people, then what?" And nobody could answer. Now, the only thing I can offer them is that there is no necessary answer. But that is the next task.

The task beyond that is to go to Budapest and try to talk to local governments about how to encourage entrepreneurship, and the reason for doing that is without a sound economic base you do not have the surplus in your economy to fully educate your workforce. By the same token, you never get that surplus unless you do fully educate your workforce. It is a vicious cycle which we would be not serving ourselves well if we broke.

The first thing I wanted to do was to give you some information about how we compare. I say that because when I read about the national position on testing, which is what one finds in newspaper editorials, it is always about how badly we do. I think to compare national systems with each other is nonsense on the whole because each one is emphasizing different parts of its product, and that is particularly true in a place like the United States and to a lesser extent Germany and a couple of others where there is significant variation even in the types of workforce being educated in some parts of the countries.

I did include in the beginning of my testimony, which I won't repeat here, four or five tables and charts which will show you that we don't do that badly. I will say this: If you wish to control for the number of students tested, and on some tests, for example, the pool from which samples are drawn in West Germany is only 9 percent of their student population from gymnasiums. We sample from all 75 percent of our 17-year-olds currently in school. Whether you control for that or not and whether you choose your test correctly, you can make almost any point you want about how well we do versus somebody else. For once I wanted to make a good point, so that information is there.

It is not that I believe that we do awfully well in every aspect of our educational systems in the various States, but I will come to the particular problems in a minute.

The second misconception I wanted to clarify was the extent to which national testing goes on overseas. We are often told we need to do it like some other country does. Very frequently, Great Britain is an example given. I would be embarrassed, given the extent to which the British educate their kids at all levels of education, to make that comparison. If it makes any sense at all to compare outcomes of educational systems, there are only one, maybe two countries that are undeniably at the bottom of the league, and Great Britain is the one that is undeniably at the bottom. That is even true, incidentally, with respect to very well-trained individuals but for an odd set of reasons.

Countries overseas don't go out and nationally test everybody. When the Brits test 7-year-olds, they are doing that kind of thing really for the first time. They have things like O levels and A levels and GCEs and stuff like that, and we tend to compare our high school graduates with people who do very well on those exams. But people forget that on A levels you can get not just an A or B, but also a C or a D, and many people do. Indeed, most do.

The same is true with the three versions of the French Baccalaureate. Roughly the same portion of people take the French Baccalaureate as get a New York State Regent's diploma among our high school graduates, or among our high school age students altogether, even including the dropouts.

So basically, when you start looking across countries, and the size of their countries versus some of our States, it turns out that very often they are testing sometimes in about the same way we test, about the same proportion of students on outcomes-based measures. And it is fair to say that some of these exams I have just described in other countries are outcomes-based measures.

If you look at the Japanese there is even a more interesting comparison. We hear that they have an end-of-high-school exam. We have to be correct about that. That is a beginning of college exam. Every high school does not take that exam. Only 40 percent of high school students are going to go to the type of college that requires that exam. Two hundred and fifty percent of Japanese students take this exam that you hear about them committing suicide over.

As you go into high school, everybody does take an exam, but I won't describe the system. I describe it in the written testimony. But it is not one exam. There are prefecture exams for prefecture high schools. There are city exams for city high schools. Some high schools have no exams.

Further, when you look at the Japanese college entrance exam it makes testing in the United States look very good. There are people who score very high, for example, on the English exam who know points of grammar that nobody in this room knows who cannot speak English. It is, basically, these are the building blocks in their most simple-minded sense.

I am very encouraged, incidentally, coming from New York State because New York State has done a lot of work on serious outcomes-based testing. This leads me to look at what kinds of national testing we do in the United States. We test more people at every grade level than any other country with three versions of the same test. The three versions are called the Iowa, California and Stanford Achievement Tests.

One of the reasons we do this has already been alluded to. In order to receive our what is now Chapter 1 money and benefits from various other Federal programs, and to decide when students are no longer eligible for those services, we have to take these normed reference tests. You folks make those rules, we don't.

Our particular district has found sneaky ways, using the New York State tests, outcomes-based tests which can be nationally normed, incidentally. These things are not at odds with each other. It is just that you don't want to pay much attention to that norming. But since you require it and New York State has taken the effort to norm a few of its outcomes-based tests, we use that to get our Chapter 1 reading money, and then we have to go back to the New York State money, the equivalent of Chapter 1 money, and put that money into our math programs.

Now, there is a lot of criticism of achievement tests and I share them. I think that a major underlying theme all through what is being said is that we are beginning to learn how to test outcomes pretty well and achievement tests or that formatting doesn't do it. But the criticisms are there, and there are two things you need to understand about those tests.

One is the Lake Woebegon effect. If it tells you anything at all, it tells you that the average, say, seventh grader knows more things that are put on multiple-choice tests than the average seventh grader 12 years ago, 7 years ago, 5 years ago. If you are bothered by the Lake Woebegon effect, just renorm the test to this year's population.

Secondly, what people never realize about achievement tests—and this is in response to a question asked earlier—is that there is very rich material there to help teachers. And again, the only use I generally have for tests is for diagnosis. It tells me what a student knows in terms of what they are supposed to know and why they don't know it if they don't. Every wrong answer is on an achievement test question for a reason and patterns of reasons accumulate, and it is very important for a teacher to know whether or not a person gets a particular score because they didn't finish reading the section or don't understand the use of semicolons. That information is out there right now and we don't get it, and, obviously, then we can't use it.

Now I am not in favor of going out and producing another test. The experience in other countries, and I think particularly of England and of Northern Ireland, is that producing another test doesn't change anything in terms of distribution of educational outcomes. On even very bad tests you know who is coming out at the bottom now, and if you have new tests three years from now, it will be the same people from the same places with the same demographics behind them. This is known.

Also, there is not a great deal of disagreement among roughly a million classroom teachers about what people should know. I don't find that when I go around. I even find general agreement across countries about what kids should know roughly on average at an age level. You know, if they are being well educated.

To act in a vacuum, as if a million teachers have never thought about this, is really duplicating an effort that is already going on on a massive scale. States have done it; and, again, New York has thought about this, with teachers, with testers, and there are even some tests out there. And New York is not the only State. It is not a mystery.

Now I am concerned about having more tests for another reason, which I alluded to in describing my trip to England. If there is a test that will be used, and particularly if there is another test, you can't necessarily, either as the producer of the test or as somebody advocating educating all of our students to their full potential, determine what use that test is going to be put to. It is in the hands of the user, and there are many hundreds and thousands of users in the United States.

Now, I do have some suggestions about what can be done. It seems to me that talking about testing, as contrasted to making our assessments, which are many and getting better, even more accurate, is putting the cart before the horse.

The things I would do, and these will involve both items I think that we have to keep in mind and some specific recommendations.

One would be to support programs for children, families, agencies, preschool and elementary schools that will help all children, say, at age 8, age 6, age 3—pick your age—all of them be able and anxious to learn, and schools be able to teach children of diverse cultural backgrounds effectively. It is not just the child's responsibility to be able to learn. The schools must be ready to teach this very diverse set of children that we have in this country.

You have to understand that no country has ever tried to educate such a diverse set of students since the days of the Roman Empire, and, obviously, it is something that everybody in this room seems to value.

The second thing to understand about that is that when we do this, and we have to educate all our children, even children of illegal immigrants when their parents are in jail—that is since 1984—it is going to cost more, not the same money, to produce the same outcome with that very heterogeneous group than would be, say, true in Germany. In Germany, except around Frankfurt, there are two things that are true of a German classroom: they are all white and they all speak German. Try that even in Ithaca, New York. We do ESL and transitional bilingual in 51 languages.

If a class of 30 or 35, or even 40 or 45, in the Japanese case works effectively for most students, it is not true here, if you want those diverse students in the same classroom. The best teachers I have ever met can handle the low twenties with that kind of classroom. All of the studies of class size have been done after you have imposed this very funny tracking system that has implicitly been part of American education historically onto the schools, and you are dealing with reasonably homogeneous classes. So many schools are self-consciously moving away from that. You cannot use the old studies to justify large class sizes.

You also need to understand that teachers work about the same hours as other professionals. The teacher is the only professional I know whose real wages did not go up between 1972 and 1988.

I, obviously, have in here recommend continued support for Chapter 1 and Head Start.

Finally, provide Federal support for technical apprenticing after high school. Only in the United States and a few other countries do we keep all of our kids in the same school. These tests that are given to some kids earlier, and they are not always national tests—in Germany they are state-based, sometimes grades in school and fourth grade matter in some states in Germany as to whether you go to gymnasium or not. But what they all wind up doing, other than a very few countries, particularly the United States, is physically putting the kids in different places at an age where the kid can't make that choice intelligently for themselves or understand their consequences.

It has less effect there, because the ratio of blue collar salaries to white collar salaries is higher in most other countries than in the United States. Alternatively put, income inequality is larger in the United States than most other industrial societies and is the only one in which it has been increasing the last 10 years. So, if you did that with our children, you are making a decision of immense consequence just in terms of how—what kind of food they can buy.

We don't do that, though. We keep them all in the high school, even though we covertly move them out from the different little groups. But we try. But, obviously, they don't all get through, and the problem we have is, in my eyes, different. The problem we have is that we do have 30 percent of our students in the system who are not prepared for the workplace or further education. It is not that everybody is failing. It is not that we have a horrible system. It is not that most people aren't making conscientious efforts to teach all students effectively. What do we do about the 30 percent? And basically, I have been telling you, if you think it doesn't cost money, I think you are very mistaken.

I also think you have to own up to what the problem is and not think that a national test or people sitting here, particularly people like me, can do anything about it until we get out into Ithacas, Schenectadys, San Diegos, wherever.

There are a couple of particular things that would help, though. The German example is very good in this sense. What the Germans do with respect to those kids they do separate out is they work through partnerships with businesses, as you know, but in the process of technical education, which I would advocate beyond high school as an entitlement, they continue academic education. Part of their apprenticeship is not only in classroom work in these skills, you could call it essential work skills, but continual education in reading, writing, math, et cetera. There is no reason we have to do that for our 16-year-olds going into high school voc-ed in New York. It ought to be true for 18, 19, 20, 21-year-olds as well.

Now, how do you get them to stay there? Actually SCANS can be a very big help on that. I have seen a number of the SCANS workplace scenarios that are very well done. What I plan to do with that in every school district I get into is take those workplace scenarios in to business people and the school board members and say, "Look. The first thing that has to happen is let's take middle school kids because that is when aspirations and expectations diverge, and let's show them that what they are learning in sixth, seventh and eighth grade is exactly what they have to know if they want a job." And we can do that by using the SCANS workplace profiles. For that purpose, they are extremely well done.

The employers then around you are alerted that in fact your school is teaching those skills, and the kids know that those skills have to be mastered to get any kind of job other than flipping hamburgers.

One last point on this. I would like our business people, and we are working on this in Ithaca, to use high school grades. The biggest resistance I have found there and elsewhere is a general view they have of what I would call the maturity of the high school graduate, not their competency. You will take the same kid age 18-19 with the same transcript, an employer won't hire them. They will flip hamburgers for 10 years—same transcript, no intervening training—they will hire them.

There is an age bias out there too with respect to young people being employed in our workforce. Employers don't believe they are skillful. Involving the employers in individual school districts with SCANS-type programs will educate employers to the fact that they do have skills. If a national test comes along in that area, I will do my best to make sure it never is given in Ithaca.

Thank you.

[The prepared statement of Dr. E. W. Kelley follows:]

Can National Tests Affect the Quality of Education

Testimony before the Subcommittee on Elementary and Secondary Education of the House Committee on Education and Labor. March 14, 1991

E W Kelley
Cornell University
Ithaca, N Y 14853

[illegible]

BEST COPY AVAILABLE

Table 3-1. Rank Order of Achievement Test Scores by Subject and Population Group, for Selected Developed Countri and Subjects: 1970-71

Countries	Science		Reading comprehension		Literature		Civic education	
	POP I	POP II	POP I	POP II	POP I	POP II	POP I	P
Total countries with test results	14	13	12	12	8	8	7	
Austria	3	2	(NA)	(NA)	(NA)	(NA)	(NA)	(N
Belgium (Fr.)	13	11	6	4	6	5	(NA)	(N
Belgium (Fl.)	11	8	9	6	8	7	(NA)	(N
England	5	4	3	2	2	3	(NA)	(N
Finland	8	9	5	8	3	6	4	
France	10	10	(NA)	(NA)	(NA)	(NA)	(NA)	(N
Germany	2	11	(NA)	(NA)	(NA)	(NA)	2	
Hungary	7	5	11	10	(NA)	(NA)	(NA)	(N
Ireland	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	7	
Israel	(NA)	(NA)	8	7	(NA)	(NA)	(NA)	(N
Italy	12	13	10	12	7	8	(NA)	(N
Netherlands	4	(NA)	14	9	(NA)	(NA)	5	
New Zealand	1	1	1	11	1	2	1	
Romania	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(N
Scotland	5	6	2	3	(NA)	(NA)	(NA)	(N
Sweden	9	3	7	5	4	4	3	
United States	14	7	12	1	6	1	6	

Note: POP I refers to all seniors enrolled in an academic curricula and POP II refers to a subset for these students as defined in the text.

Source: Richard M. Wolf, *Achievement in America*, (Teachers College Press, Teachers College, Columbia Univ., New York, London, 1977).

" The U.S. has the lowest attrition rate, with Sweden and Japan close to the U.S. The other developed countries have much higher attrition rates. Thus, the very poor showing of the U.S. on POP I score rankings is at least partly due to differential attrition. The data underlying the ranking for POP II incorporate an adjustment for the effects of differential attrition on test scores. However, the adjustment procedure is biased in favor of countries with low attrition -i.e., the U.S. -and therefore overadjusts for the bias. Thus the true relative ranking of the U.S. is somewhere in between the results for POP I and POP II. "

U.S. Census 1985

Enrollment Rates for Students Aged 15–18 and 20–24

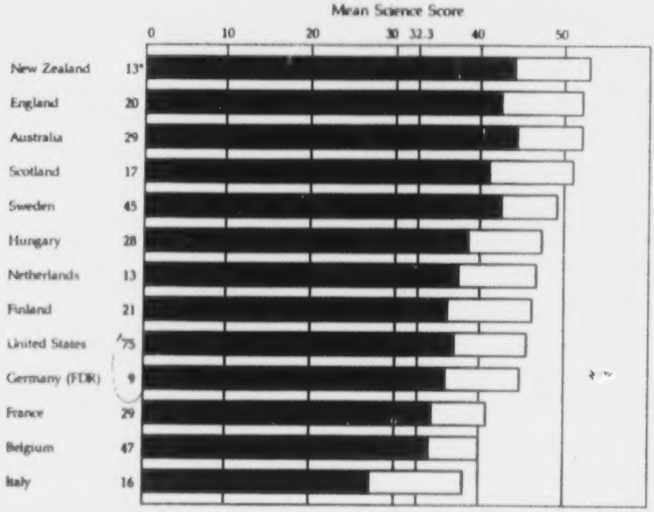
	Age 15–18			Age 20–24		
Country	*Year*	*Percent Enrolled*	*Ranking*	*Year*	*Percent Enrolled*	*Ranking*
Austria	1969	31.9%	14	1972	14.8%	12
Belgium	1966	54.2	7	1971	20.5	6
Canada	1970	78.1	2	1970	28.5	2
Denmark	1970	51.7	9	1972	23.7	3
Finland	1967	47.4	10	1972	13.8	13
France	1970	54.3	6	1972	17.0	10
Germany (FDR)	1969	30.5	16	1972	17.1	9
Ireland	1971	47.0	11	1970	12.2	14
Italy	1966	30.8	15	1972	20.1	7
Luxembourg	1970	37.4	13	1972	1.9	16
Netherlands	1970	52.5	8	1971	21.2	5
Norway	1970	68.9	3	1972	19.0	8
Sweden	1972	68.1	4	1972	22.4	4
Switzerland	1970	62.6	5	1972	10.3	15
United Kingdom	1970	39.4	12	1972	15.0	11
United States	1970	82.9	1	1972	51.5	1

Source: Rates for 15- to 18-year-olds adapted from UNESCO (1974, Tables 15 and 16); rates for 20- to 24-year-olds adapted from UNESCO (1975, Table 3-2).

The United States enrolls a higher percentage of young people in secondary school and college than does any other Western nation.

From:
Michael Kirst
Who Controls our Schools.

An International Comparison of Test Scores in Science of Students in Last Year of Secondary School



*Percentage of total age group tested.

Source: L. C. Comber and John P. Keeves, *Science Education in Nineteen Countries* (Stockholm and New York: Almquist and Wiksell and Wiley-Halsted Press, 1973).

■ Overall Score ■ Top 9% ■ Top 5% □ Top 1%

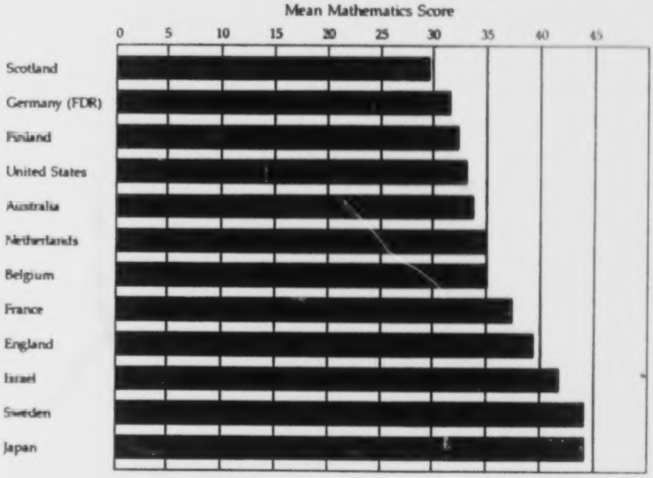
International comparisons involving the United States are skewed by the egalitarianism of the U.S. system. When scores compared are adjusted to reflect comparable proportions of the relevant age group, U.S. scores are similar to those of other countries.

From:

Michael Kirst

Who Controls our Schools

An International Comparison of Test Scores in Mathematics of Students in Last Year of Secondary School



Source: Torsten Husén, ed., ***International Study of Achievement in Mathematics: A Comparison Between Twelve Countries*** (New York: Wiley, 1967).

■ Average Mathematics Test Score for Equal Proportion of Age Groups
■ Average Mathematics Test Score for Country

When scores of all students studying mathematics in the terminal year of secondary school are compared, the United States ranks last, in part because U.S. schools enroll a much higher percentage of the relevant age group in mathematics (18 percent as against 5 percent in some European countries). When scores of equal proportions of the age group are compared, U.S. students perform about as well as students in Australia, Belgium, and West Germany.

From:
Michael Kirst
Who Controls our Schools

THE LESSONS FROM ABROAD:

Japan has national exams [illegible] exit from high schools. High schools are operated by different public sector authorities and private [illegible] ranked largely [illegible] the results [illegible] high school graduates [illegible]. These schools range [illegible] from 100% [illegible] (Rohlen, 1983).

Type of School	Percent to University	Percent to College
Prefectural academic	72	14
City (some) academic	12	28
City commercial	[illegible]	[illegible]
City night (part-time)	[illegible]	[illegible]

Most of the students who will get into Tokyo University go to the prefectural academic high schools. A few go to city academic schools. An increasing number go to rather new, elite, private schools.

From: Thomas P. Rohlen
Japan's High Schools

BEST COPY AVAILABLE

Exams for both high school and college entrance are exclusively in a multiple choice, timed format and are essentially factually oriented. A student can know points of English grammar we are almost never exposed to, but yet not be able to converse in English at even a primitive level. Many Japanese are fully aware that their national exam system does not always measure even the skills, let alone more complex problem solving abilities that are most valuable to both student and potential employer. Yet there is a strong pressure (for reasons not given here) on some students to attain a high level of skill at rote and some complex academic skills. As can be seen in the accompanying diagram only a minority of Japanese students age about 18 take their "national" exam used to screen into universities. No Japanese exam is designed to assess their education of all students or give the country a report card. The same is true in France where the baccalaureate is still taken by fewer than 50% of students in their late teens. Actually the baccalaureate is a number of different, overlapping streams of courses with curriculum fixed hours of instruction and final exams mandated nationally. Grades are given to students as a consequence of the exam in each subject that forms part of the particular baccalaureate they take.

British A level exams are pegged at around the same level as the baccalaureate, but allow greater students choice in combining subjects. While there are three major variations of the baccalaureate, British students can mix the A-level exams they take in many ways. About half the percentage of students in Britain take A-levels when compared to French students taking some baccalaureate. This is consistent with an almost universal observation: if comparing the products of whole educational systems made any sense at all, Britain is probably at the bottom of the league of industrialized democracies. Most students who finish secondary school never take more than a few O-level exams; these exams can be passed by most minimally competent U.S. high

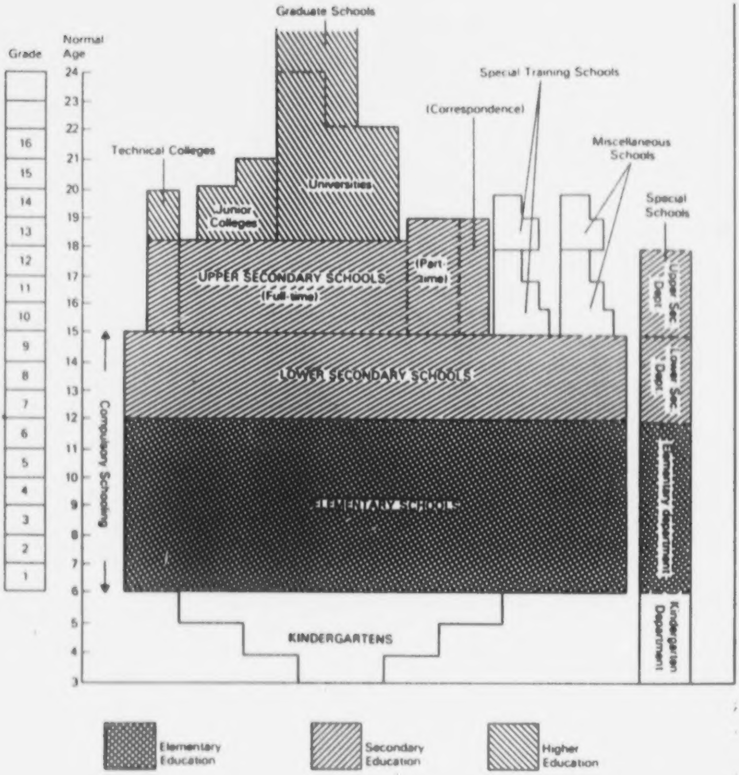

Graduate Schools
Grade
Normal Age
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
Technical Colleges
Junior Colleges
Universities
UPPER SECONDARY SCHOOLS (Full-time)
(Part-time)
(Correspondence)
Special Training Schools
Miscellaneous Schools
Special Schools
Upper Sec. Dept.
Lower Sec. Dept.
Elementary department
Kindergarten Department
LOWER SECONDARY SCHOOLS
Compulsory Schooling
ELEMENTARY SCHOOLS
KINDERGARTENS
Elementary Education
Secondary Education
Higher Education

CHART I. Organisation of the Japanese Education System. From *Basic Facts and Figures about the Educational System in Japan* (National Institute for Educational Research, Tokyo, December 1983).

school graduates.

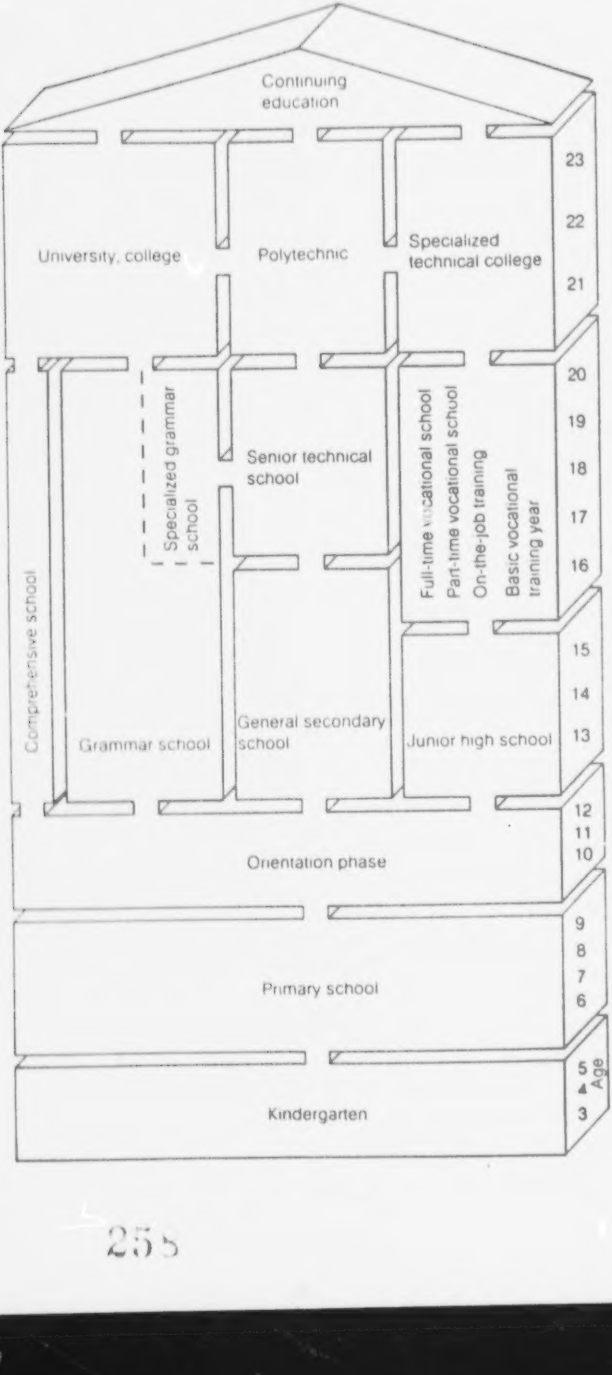
The level at which the British and French exams are pitched is about that of the average end of first year exam at a U.S. academic four year college (There is some variation across these colleges.) While the typical recipient of a baccalaureate or a British student successful at the A-level math would probably know more mathematics than the U.S., four year college freshmen, that subset of freshmen who will major in math and science, the fair comparison group, will know about as much. However, there would be twice the proportion of French and U.S. students involved in the comparison.

Like Britain and France, Germany has no national assessment or any national test that almost all students take. Like the United States, control over educational instruction and evaluation lies principally at the state (laender) and local level. As in the United States, frequent professional contact across states and a national media result in great commonality of curricula and materials examined by state or local educational authorities.

German states vary in the age [12-14] at which they first formally separate students into more or less academic tracts; no uniform assessment process is used across the country. Fourth year grades, some exams and teacher judgement (rarely parental judgement, unlike the United States) are all possible elements of a decision to track a child to gymnasium. Historically the possibility of movement into white collar professions is largely determined at this point, the child can hardly be said to be an active and informed participant in choosing his or her future. Yet the sifting process is largely not resisted for two reasons. Study after study shows that students around the ages of 12-14 absorb the expectations (rather than aspirations) of those around them, including, critically, those of their peers. These expectations are largely based on the historical and inexpensive information provided by parental education, job status and income

Germany

Structure of the Education System


Continuing education
University, college
Polytechnic
Specialized technical college
Specialized grammar school
Senior technical school
Full-time vocational school
Part-time vocational school
On-the-job training
Basic vocational training year
Comprehensive school
Grammar school
General secondary school
Junior high school
Orientation phase
Primary school
Kindergarten
Age
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3

(for the U.S. add race and language variation). Second the status and financial penalties of not being in a white collar profession is not nearly as great in Germany as it is in the U.S. The Germans have an excellent apprenticeship system run by industry and state governments together. Part of the program is school based. Products are largely literate workers with skills to work productively in a particular industry. These workers are well paid (almost always as much or more than their U.S. counterparts even on the occasions when U.S. workers are as or more productive). Inequality of earned income across white collar, blue collar workers, managers, etc., is lower in Germany than the United States; only in the United States have all measures of income inequality increased in the last ten years. Of course, in Japan, inequality of earned income is the lowest of the countries we are considering and the percent of the economy totally in the private sector is highest.

None of these countries conduct internal assessments like our NAEP. All are involved in multinational survey comparisons, many of which involve the United States as well. The Council of Europe in Strasburg probably has the best information about national and even local educational practices and tested performances. None of these countries has a single test that all students take. The Japanese end-of-high school exam is not taken to graduate, rather it is taken to get into college. Almost all of the high school graduates in France take some form of the baccalaureate. However that number represents not much higher a proportion of students that those in New York State who receive regents diplomas.

NATIONAL AND STATE TESTING IN THE UNITED STATES

Schools in the United States already give almost national tests. These are the Iowa, Stanford, California achievement tests. School districts must use these or other nationally normed tests to determine eligibility for various federal programs. The achievement tests are comparable and are less oriented to rewarding rote learning than even the Japanese entrance exam. However, achievement tests evaluate each student in terms of the curriculum of no one particular school district! Further, like most of the tests we have mentioned, achievement tests are passive, bubble sheet, multiple guess, timed exams. The question of how to adequately assess the extent to which students master the outcomes we set for them is important in determining the adequacy of education, but must await the political courage to be explicit about outcomes and the development of performance based testing. The Lake Wobeggon effect of achievement tests tells us that students on average know more than students of the same age did five or seven years ago. The Lake Wobeggon effect can be eliminated by simply renorming the test. The achievement tests are given to more students than NAEP could ever sample. People are excluded from both tests for the same reasons (i.e. some type of handicapped conditions). NAEP oversampling would be unnecessary, if achievement test scores were being compared. Achievement tests contain many more questions and are arguably more closely tied to what some educators think children should know after a number of years of schooling. We could dispense with norming achievement test scores and use the tests as if they were outcomes based on a general curricula. Oddly, we do not use achievement tests to help teach. Since every wrong answer is on a multiple guess test for a reason, to ascertain patterns of wrong responses could be valuable for teaching.

Scholastic Aptitude Tests are not national tests that measure outcomes, even though they are included in the Department of Education "wall chart". SATs can only be considered as predictors of collegiate academic performance, but they are such inefficient predictors, we need not consider them farther.

Individual U.S. states are often as large as many of the countries with which we compare our educational outcomes. Our larger states probably achieve full economics of scale in evaluating student performance. State minimum competency tests do not demand sufficiently high levels of mastery in any subject. One reason the U.S. has more high school graduates than other countries is that the common denominator of graduation is a low level of accomplishment at academic subjects. Yet some tests, like many New York State regents exams require levels of performance of students that would place them toward the top of their age peers in most other countries. New York has also developed outcomes based tests which compare favorably with diagnostic performance tests given anywhere (Unfortunately these are often still in a multiple guess, bubble sheet format) (See tables 1-2; figure 1) The problem is not in the existence of ever better and more diagnostic tests, but in the fact that a number of students do not do well enough.

We can conclude that the United States tests as much as any country. (For some guidelines about testing see Figure 2). Probably some state tests are more self-consciously based on educationally desirable outcomes. Our nationally normed achievement tests are taken by a higher percent of our students each year than take any tests in any other country. No other country has NAEP on such a scale for so long. Two caveats: almost everything we do with NAEP can be done more accurately with currently given achievement tests; why give the NAEP if there is no consequent policy change at national or state levels that will effect educational outcomes at the classroom level. Information is a costly waste if it is not in a form that can be used and then is used to change outcomes. More national testing in the U.S. will have a feww ecrtain results: less instructional time and more professional testers employed.

MISIDENTIFYING THE PROBLEM

National tests are not the problem with U.S. education nor are they the solution to any problem. Our problem is not that education is overwhelmingly failing us. Rather some of us are not being well educated in school and no other provision is made to help these people be useful in the workforce (unlike Germany and Japan). Some of these people even receive high school diplomas. In the future our work force will require a distribution of skills we are currently not providing in our educational system. (Figure 3, 4).

When students are tracked out of academic education in Germany, they are simply dropped. In Germany all students have access and the later differential between their income and that of white collar workers is small(similarly in Japan). Students pay little or nothing for this training, cost is shared by federal and (mostly) state government and business in Germany. The savings occur in later tax collections and less on the job training. The apprenticeships all involve continued general training in literacy, writing and numeracy. The two most populer apprenticeship programs in Germany are automechanic and hairdressing, the same as in the United States.

The distribution of those who do not achieve high levels of educated competency in the U.S. is well known. They are diproportionately black, hispanic, and low income.(See for example Figue 5). Eighty seven percent of children under three born of black never married females live under the poverty line. We already know many of these children will not attain high levels of competency in school for a complex set of reasons none of which was initially a choice the child made. It is unfair to blame the child for events he/she did not choose and insufficient to do nothing to make that child a competent citizen with skills to enter the work force. More national testing will neither improve that or any child's chance to become successfully educated nor will the pattern of failure be changed. Indeed changing or adding to national testing can often exacerbate inequality of opportunity [illegible] achievement [illegible]

[illegible] that will ultimately change the [illegible] educational outcomes in classrooms and schools.

2. Support those programs for children, families, agencies, pre-school and elementary schools that will help all children at, say, age 8 be able and anxious to learn and schools be able to teach children from diverse cultural backgrounds effectively. Do this without imposing a monolithic culture on families. Understand that we are the most heterogeneous culture ever, with the possible exception of the Roman Empire.

3. Understand that to educate all our children (which we must since the Supreme Court formalized this practice in declaring Section 21.031 of Title 1 of the Texas Educational Code unconstitutional) will cost more than would be true in homogeneous cultures. Class sizes of 30-35 are possible if all the students are white and speak German; class sizes much above twenty are difficult for a skilled teacher, if the student population is heterogeneous.

4. Understand that teachers work the same average hours as other professionals and work at tasks most of us could not do. Teaching is the only profession whose real wages did not go up between 1972 and 1988 (Non-professional real wages went down in the same period).

5. Continue support for Chapter One and Head Start.

6. Provide federal support for technical apprenticing after high school. Only in the United Sates and a few other countries do we keep all our children in one school through age 17-18. Children

BEST COPY AVAILABLE

grow, change their interests and progressively make choices for themselves (much more than the Germans for example). We should not be making choices for youth that they should make for themselves. It is they who live with the consequences.

Be sure that increased general literacy, writing skills and numeracy are part of the program. Provide incentives for states to organize the programs and insure that it work on site at businesses as well as formal training.

Make it free or almost free. [illegible]

[illegible]

[illegible] national- [illegible] educational outcomes or their distribution in the population.

BEST COPY AVAILABLE

TABLE I

TESTING SCHEDULE

to measure high competence without norm-referencing

GRADE/TEST	READING	WRITING	MATH	SCIENCE	SOCIAL STUDIES
Kindergarten	Writing voc Dictation	Sample	Checklist		
First	Writing voc Dictation	Sample	Checklist		
Second	DRP	Sample	PEP		
Third	PEP	Sample	PEP		
Fourth	DRP	Sample	PEP	PET	
Fifth	DRP	PEP	PEP		
Sixth	PEP	C#1 Sample only	PEP		PET
Seventh	DRP	Sample	PEP		
Eighth	PCT	PCT	*TBD		PET
Ninth			RCT	RCT	
Tenth					RCT Global Studies
Eleventh	RCT	RCT			RCT US History

REGENTS - 7-8 [illegible]

(424)

* To be determined/developed

RCT - Regents Competency
PCT - Preliminary "
PEP/PET - (State normed tests) outcomes based

Some tests are normed on a state or national level after development. All are created as outcomes based.

Table 2. Readability of Prose Samples in DRP Units

34 DRP Units

Bears are big. They need a lot of food. Bears eat meat. They eat bugs. They eat berries. They eat honey. They eat fish, too. Bears feed in the spring. They feed in the summer. They feed in the fall. Bears look for food then. They hunt. They fish. They dig roots. They pick berries. They eat a lot. They grow fat. Soon, winter comes. It gets cold. It snows. But the bears don't need to go out. They don't need food. They are fat enough. They can sleep.

3rd grade IRP } 36

39 DRP Units

A bird's wings are well shaped for flight. The wing is curved. It cuts the air. This helps lift the bird. The feathers are light. But they are strong. They help make birds the best fliers. A bird can move them in many directions. Birds move their wings forward and down. Then they move them up and back. This is how they fly.

43 DRP Units

Many states are dry in summer. They get hardly any rain. Nearly all their water comes from melted snow. It is stored. It is kept in dammed up ponds and lakes. It is used during the growing season to water farms and orchards. Farmers buy the water. They are told how much they will be able to get. The amount changes each year. It depends on how snowy the winter was. A farmer needs to know how much he will receive. It allows him to decide which of several crops he ought to plant. The choice is based on how much water different crops need.

47 DRP Units

The part of a beach between high and low tide is called the middle beach. It is home to many plants and animals. But life on this middle beach is hard. There is no protection against the wash of the oncoming waves. Some animals survive by digging holes in the sand. They can stay in their homes under ground. The undertow will not pull them out to sea. They are safe.

51 DRP Units

Most creatures take care to protect their eggs. The walking stick does not. It just drops its eggs, scattering them loosely on the ground. Dozens and dozens drop at a time. As the eggs fall onto dry leaves, they sound like raindrops falling. Many of the eggs do not hatch. But enough do so that the walking sticks will not die out. They have existed on earth since before the era of the dinosaurs.

th rade IRP } 53

56 DRP Units

The people of Greece used the alphabet of the Semites. At first the Greeks wrote from right to left and left to right in alternating lines. The Greek name for this system of writing came from their words for "ox" and "turn." The method reminded them of oxen going back and forth, plowing a field. Eventually, the Greeks wrote only in one direction, as most people do now.

60 DRP Units

The ouija board is a simple rectangular piece of wood. All the letters of the alphabet are set out in a semicircle across a long edge. The ten digits and the words "yes," "no," and "goodbye" appear below. A small heart-shaped piece of wood called a planchette is mounted on casters so it can move easily on the board. When one places his fingertips lightly on the planchette, it slides around. It moves apparently without any conscious control on the part of the operator. Its pointer is supposed to spell out the answers to questions.

64 DRP Units

Wall paintings are especially vulnerable to atmospheric change. Archaeologists know this. Hence they try to discover, before opening a tomb, whether they will find murals. Special tools have been designed for this purpose. One of the most useful is a kind of camera that can be dropped into the ground before the digging starts. If the camera indicates the presence of wall art, scientists can prepare to take steps to preserve the painting as soon as it is reached.

RCT 70

73 DRP Units

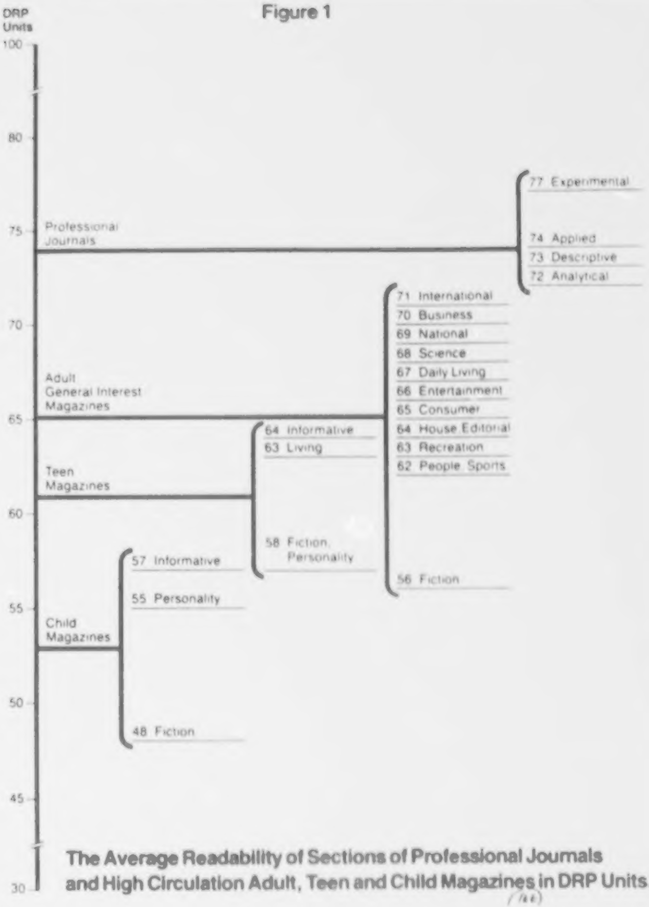
Hellenistic literature showed an interest in individual history and psychology, rather than man in general. Theophrastus' Characters, with its detailed portraits of such types as the flatterer, appeared during this time. Biography, dealing with the lives of real people, was a flourishing form. And in philosophy the emphasis was on personal conduct rather than speculation about reality.

81 DRP Units

Jefferson's preference for an agrarian society and his idealization of the independent farmer reflected a conviction that representative government required a secure and relatively prosperous economic base to function successfully. He perceived the farmer as economically independent, and thus unlikely to surrender his judgment as a citizen to the influence of demagogues. His dislike and distrust of cities derived from a conviction that urban conditions, especially for the poorer classes, forced men into such a bitter struggle for sheer self-preservation that their natural moral sense could not be relied upon to produce social harmony or to guarantee responsible citizenship.

Note: The readability calculations are based upon longer samples.

Figure 1



The Average Readability of Sections of Professional Journals and High Circulation Adult, Teen and Child Magazines in DRP Units.

Figure 2

WHAT SHOULD BOARD MEMBERS KNOW AND\OR DO ABOUT TESTING

1. Uses of Testing:
   - a. evaluation of curriculum over district
   - b. extent to which student has/has not mastered curriculum
   - c. reasons student has not mastered curriculum distractor analysis
   - d. prediction of future non-test performance

   Note: Remember: we are teaching children information, skills, and how to reason; we are not teaching chilren about tests. (unless the rest of the world makes us.)

   Testing has no automatic consequences that are not negative.

   Don't give a test unless it does a-c above. In particular, ask: when test results come in, do they tell us what we must change in teaching a particular child or in the curriculum. If no change can be inferred don't give the test.

2. Using Tests:
   - a. bringing the task into the test versus bringing the test to the task: K-1 testing.
   - b. achievement testing. Do they satisfy any of the uses given above. Standardized tests versus national norms.
   - c. alternatives to achievement testing: the problems of context, time and curriculum relevancy.
   - d. testing and state-federal funding. requirements.
   - e. curriculum based tests: role of the state
   - f. competency tests: competency at what; who decides?
   - g. predictive tests: are they worth using?

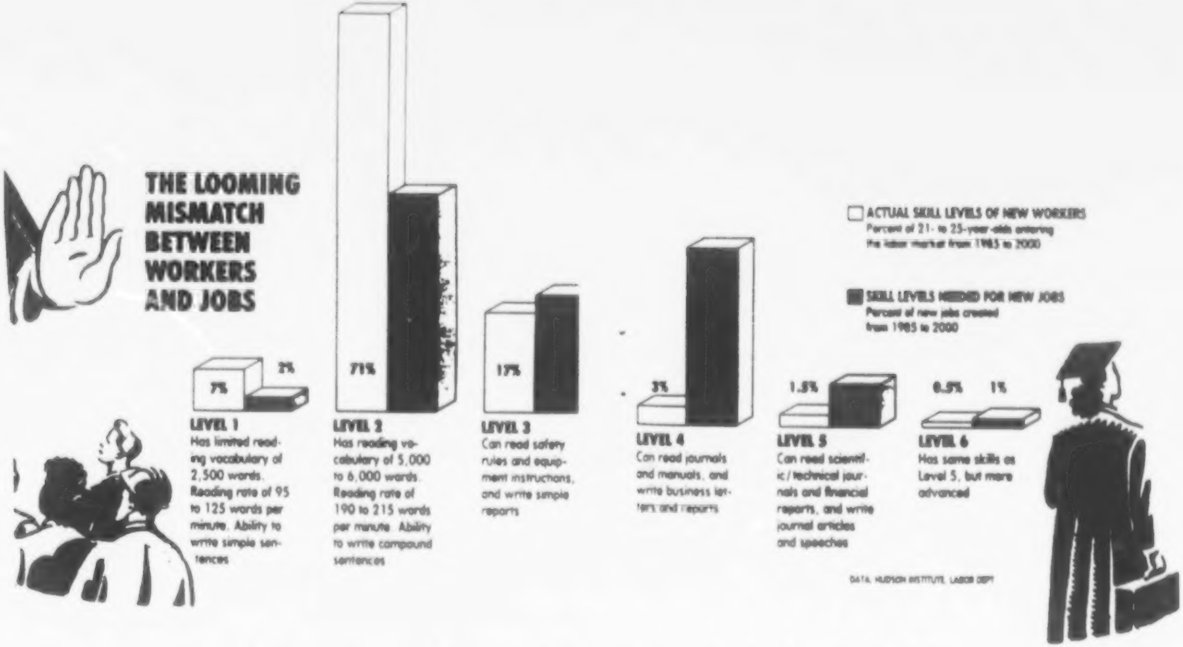
3. Report Cards:
   - a. school district's report card. CAR et. al.
   - b. the nation's report card. NAEP et. al.

4. Bias and the fear of failure: some human consequences of testing.

Ithaca City School District
400 Lake Street
Ithaca, NY 14850
607-274-2101

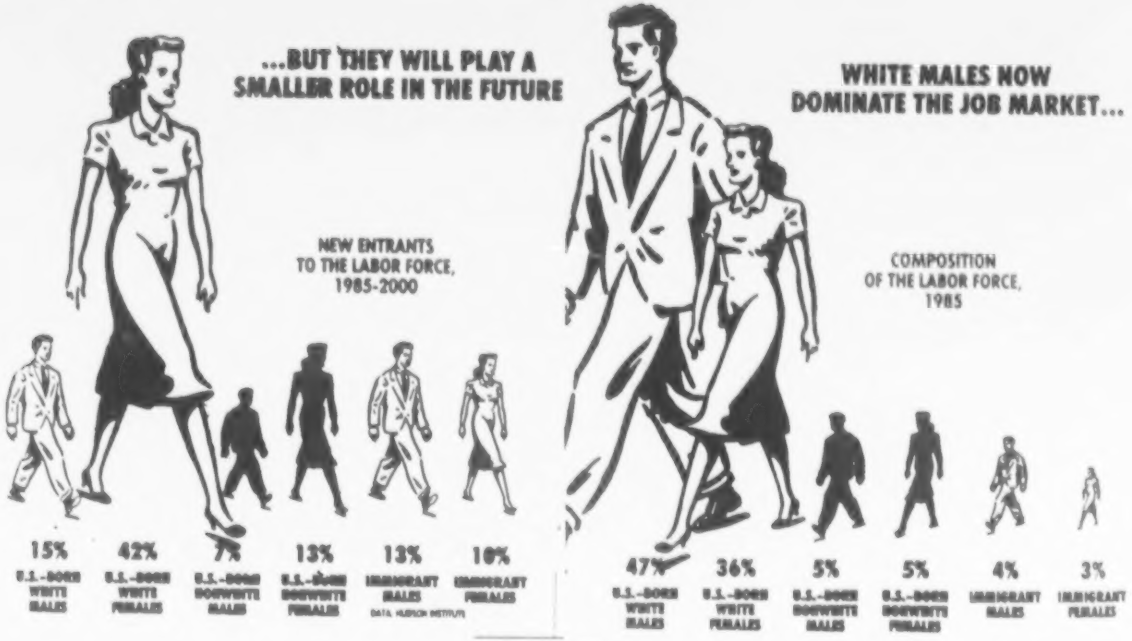
E.W. Kelley, Chairman
Board of Education
James E. Lorthridge
Superintendent of Schools

Figure 3



From *Business Week* September 19, 1988

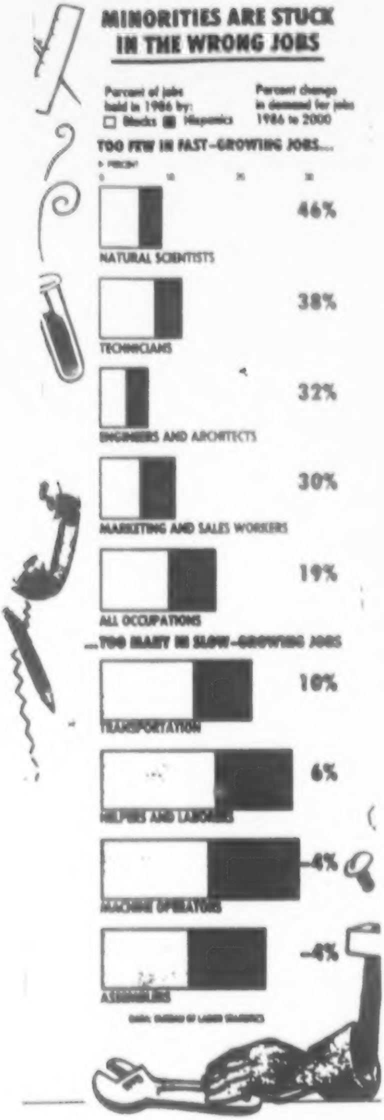
Figure 4



DATA: HUDSON INSTITUTE

From Business Week September 19, 1988

Figure 5



DATA: BUREAU OF LABOR STATISTICS

From Business Week September 19, 1988

Chairman **Kildee**. Thank you very much.

Gordon Ambach?

Mr. **Ambach**. Thank you very kindly, Mr. Chairman, Congressman Petri, Congressman Hayes, and members of the staff. You have passed your own endurance test for this morning. Performance test, I might note. You have been extremely patient and very attentive to the presentations that have been made.

I speak this afternoon, Mr. Chairman, as the Executive Director of the Council of Chief State School Officers, particularly today as the United States Representative to the General Assembly of IEA, which is the major international education assessment organization that we have in the world. And because my good colleague has been making references to outcomes-based assessments in the State of New York, which I happened to put in place, I am going to take credit for having been a New Yorker and the commissioner who was involved in that State, which does have, as a matter of fact, the most extensive assessment system in the United States, both individual exams and performance or program assessments.

Mr. Chairman, I commend you strongly for this hearing. You are smoking out the issues, and that is extremely important here. We have to be looking at what are the purposes of tests, what are different types of tests, and I would submit that already you might reach these conclusions from what you have heard yesterday and today:

One, there must be a unique United States solution to the issue of national or nationwide testing. We cannot simply borrow from somebody else what they have done because it has to match onto our own national organization, nationwide organization for education and its unique characteristics.

The second thing I think is that the real issue before you is what kind of nationwide testing or assessment system we should have. It is not really the issue of whether there should be a single test or not a single test. It is much more complicated than that. It is the question of how you put the pieces together in a nationwide assessment system.

And it has also been clear I think from the presentations that we have oodles of testing and assessment at the local level, at the school district level, school levels, State level, national level, even at the international level. A critical issue, however, is how one connects up those different systems, tests and systems, so that there is a common currency of information about the results, and that it is possible to be able to relate information about testing in Iowa or testing in Idaho with information about testing someplace else without necessarily having to test every single student in order to be able to do that.

Now, you requested, Mr. Chairman, that I speak particularly to the issue of the international dimension of national testing, and I would point out in simple form virtually every speaker that has preceded me has made the case for it, and the comments and questions that have come from the members of the committee have made the case. Just think through this morning at how many points along the way someone has made a reference to a cross national comparison of education, someone has made a reference to what rank ordering there is or made a reference to experiences in one country as against another, and that is precisely the point that I bring before you. But, as you are thinking about what might be done with national testing or national assessment, keep in mind this international dimension.

From both the understanding that we see and the misunderstanding that we have across the country, it is absolutely critical for us to be investing more in careful analysis and study of what is happening across the world. Not just in the scorecards, which are the popular public media items that demonstrate where the U.S. is related to other countries, but in a much more profound sense than that. What do we learn about the type of instruction, the nature of content? Who is involved? Is there screening in the system by way of where students are placed? What kind of instructional approaches are being given? What are the overall system characteristics? Those are the things we need to know.

Now, for about 35 years, there have been a fair number of international assessments, and the value of them has first and foremost been on trying to understand what are other systems like; and, secondly, to try to get some kind of common formatting for being able to assess pupil achievement so that you could relate what the system is like to what the achievement is, and it might help to inform us as to whether we could change or make policy for education in our country and other countries on the basis of that information.

The fact is we have a worldwide natural laboratory of different educational approaches, and whether you are talking about the issue of vouchers or choice, or the organization of schools, or 240 days or 160 days of schooling, and so on, you can if you are careful about it, as my colleagues pointed out, extremely careful about it, undertake to look at what happens where those practices have been put in place in an attempt to try to get some sense of what the result has been.

Does it automatically lead you to a conclusion as to what we should do in the United States? Of course not. But it is a natural laboratory that must be used. That is the value of the international dimension.

The second major point that I would like to make for you is that it is critical in doing international studies to have reliability, validity, fairness, an even playing field, if you will, as best you can, by way of making these studies. Within the last three years, there has been created in the United States a new board, which is the United States Board for Comparative International Studies. This document of framework and principles has been produced by that Board. It is under the National Academy of Sciences. I won't go into any of the principles or the framework, but it is an important guidepost for us all here in the States and in other countries to be using in terms of constructing these studies.

The third point that I want to make has to do with what is up for this next decade. Already on the drawing boards, in fact, in progress and with the support of both the National Science Foundation and the Department of Education in this country, Canadians with strong support is probably the largest single international study of education that has ever been done. It is called TIMSS, or the Third International Mathematics and Science Study.

In 1993-94 and in 1997-98 there will be a very extensive study of those subjects with three populations—nines, thirteens, and seventeens—and done in such a fashion that there will probably be some 40 nations that will participate. The emphasis is not going to be on just what is the achievement level, but rather on the study of the systems. In the jargon, what is now being called the opportunity to learn. What does a student have an opportunity to learn in each of these systems, and then what are the results?

I would urge that even though this is not an authorizing issue that the members of this committee be attentive to the appropriations aspect of this in order that we have sufficient funds to carry that study through.

And there are additional studies: in reading literacy, one in early childhood development and education, and a third one in the use of computers in schools, which are under way in the IEA. So, at the end of the decade there will be a significant study in how it is that children in various countries learn a language other than their native language, something which I think could help to inform us in the States greatly.

My last and final point has to do with how it is that we directly connect these international dimensions with our national testing or nationwide assessment system. I had the privilege just a week ago of testifying on the Senate side, and I testified to the effect that what we need is not a national test, but what we need is a nationwide testing system. Some aspects of it are what we call program assessment: sampling of students, a relatively low cost activity which tells you whether the system is producing; and you also need individual exams or tests which tell you whether individual children are learning toward particular goals.

What we have to think about is how to connect those two different types of tests in a national and in an international sense to common standards, and you heard that over and over this morning. Common standards against which we design those tests, so that there can be a connection. And so think in terms of what framework we need, what apparatus. I happen to believe it is a national board for purposes of establishing these kinds of standards for our assessments as related to the international dimension as well as the national.

One last point by way of these connections has to do with NAEP. The authority for State-by-State ends, unless you change it, in 1992. Many of the States which I directly represent are extremely interested in thinking down track, to want to continue to have the State-by-State option, and they are also, perhaps to your surprise, thinking about wanting to do a subsample of the IEA TIMSS study so that the State might even be able to relate its results to the international study.

The point here is that your early decisions on NAEP, the authority for State-by-State and the continuation of that, are extremely critical right now for the several States as they think about their own State assessment systems, how they are linked with the national, and how they link with the potential international studies down track.

It's an immense agenda that you have before you, Mr. Chairman, and members of the committee, but it is an extraordinarily important one. It is one on which it requires the kind of care which you have heard expressed over and over again here this morning, and we would like in the future to help you in any way we possibly can to reach the right conclusions.

Thank you very kindly.

[The prepared statement of Gordon M. Ambach follows:]

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

INTERNATIONAL DIMENSIONS OF NATIONAL TESTING

STATEMENT BY GORDON AMBACH
EXECUTIVE DIRECTOR, COUNCIL OF CHIEF STATE SCHOOL OFFICERS

BEFORE THE
SUBCOMMITTEE ON ELEMENTARY, SECONDARY AND VOCATIONAL EDUCATION
COMMITTEE ON EDUCATION AND LABOR
U.S. HOUSE OF REPRESENTATIVES

MARCH 14, 1991

Mr. Chairman, Members of the Subcommittee, and Staff Members of the Subcommittee. Thank you for the opportunity to testify at this important hearing on "National Testing: Pros and Cons." At your request, I will focus my remarks on the international dimensions of national testing. I speak as Executive Director of the Council of Chief State School Officers and as the United States Representative to the General Assembly of the International Association for the Evaluation of Educational Achievement (IEA), the most significant international organization for international studies.

I have four points to make. First, the value of international comparative studies; second, the necessity for a comprehensive framework and principles for conducting international studies; third, the importance of United States' support for IEA studies planned through the decade of the 1990s; and fourth, the relationship between international studies and nationwide assessment in the United States.

(271)

BEST COPY AVAILABLE

1. The Value of International Comparative Education Studies. The value of international comparative education studies includes the following: a) they provide us with an understanding of the differences in curriculum, instructional practices and system structures in the various nations; b) they provide measurement of student achievement in different nations in a format which makes comparisons of results possible; and c) they help us to understand the reasons for the differences in student performance in the various nations. For the United States, with the recent establishment of national education goals, international studies have a new importance in measuring comparative progress, particularly in mathematics and science, toward our goals.

Up until the 1980s, most students and parents thought about student performance in comparison with their neighbors or, perhaps, students in other schools within the school district. As one travels now across the United States talking with students and with parents, their concern is about performance related not only to the neighborhood or community but to student performance in the state, the nation, and particularly to students in other parts of the world. Our students and their parents recognize their future is in an international society and their association with peers around the globe will depend upon the education they have compared to that of students in Sweden, Japan, Singapore, Germany, Nigeria or other countries. They recognize, furthermore, their future employment is in an economy which knows no national boundaries.

The common public notion of international studies of education is typically the view of a graph in the newspaper or on television which rank orders countries according to an aggregate measure of student achievement on a particular math or science test. The assumption too often is that the only purpose of the comparison or the study has been to see who is winning the "education race."

In fact, international studies of education first began with the purpose of learning more about the nature and type of education provided in the several countries so that there could be an understanding of the varied opportunities to learn afforded students in different countries. In order to judge what results yielded from that opportunity to learn, achievement testing was coupled with these analyses of educational programs. Over a period of three decades, a series of studies have reported both a description and analysis of educational programs in different countries and the student achievement results related to those programs.

In the last 15 years, as we have become more and more concerned about the effectiveness of schooling in the United States, we have turned greater attention to the analysis of educational performance in other countries. The result of these studies has brought to light relative achievment levels of American students and, perhaps more important, drawn attention to significant differences in the characteristics of schooling, organization of schools, and the content of instruction in different countries. The variety of approaches to education provides us with a "natural" world-wide laboratory to examine different educational techniques and associated results of student achievement. We need to expand our capacity for such international comparative studies so that we may learn more and more about practices in other countries and the ways in which they may inform policymaking for education in the United States.

2. Frameworks and Principles for International Comparative Studies. International comparative studies must be undertaken with validity, reliability and fairness. Although often difficult, it is extraordinarily important to assure examination of both practices and results is done with careful design for good sampling of student population, valid and reliable testing, and so that there are fair comparisons made where there is considerable variation in instruction or the curriculum. Three years ago through the support of the U.S. Department of Education and the National Science Foundation, the National Academy of Sciences, National Research Council formed a new board, the United States Board for International Comparative Education Studies (BICES), for the purpose of providing guidance to United States funding agencies and to researchers in the U.S. and across the globe on the conduct of international education studies. The Board has issued an important report, "A Framework and Principles for International Comparative Studies in Education," attached to my testimony. There is not time here to review the report, but I would urge your consideration of it. This report has been extremely well received both in the U.S. and abroad. It provides the groundrules for United States' participation in studies in a manner to assure the funding agencies, Members of the Congress and the public of credible results from investments in studies.

3. Support for IEA International Comparative Studies during the Decade of the 1990s. IEA studies have been undertaken for 25 years. The IEA has launched the largest single international study of education ever conducted. In 1993-94 and again in 1997-98, there will be a major study of mathematics and sciences, the Third International Mathematics and

Science Study (TIMSS). Nearly 40 nations will participate in this study with sample populations of 9, 13, and 17 year olds being tested. Extensive reviews of their "opportunities to learn" -- the curriculum, instructional and school patterns in the different countries -- will accompany the testing. The United States Department of Education, National Center for Educational Statistics, and the National Science Foundation have already made significant commitments for the implementation of this study. Design work is proceeding under guidance of the IEA headquarters at The Hague and the International Coordinating Center for the study at the University of British Columbia in Vancouver. The design of this study has been significantly influenced by the United States in order that the study will yield significant information on educational programs and results related to progress on National Goal #4, Mathematics and Science.

IEA has other studies underway, including those on early childhood development and education, directed by David Weikart of High Scope Educational Research Foundation in Ypsilanti, Michigan. There are also studies of reading literacy and of the use of computers in education. Still other studies will be launched at the latter part of the decade, probably one of learning languages other than a native language of each country.

I urge your strong support for financing these international comparative studies. The general authority of the U.S. Department of Education and the National Science Foundation is in place; the key is providing resources timely for preparation of state of the art assessments, an effective cross-national cooperation needed for implementation and analysis of the results.

4. International Studies Related to National Testing. The focus of the hearing today is on national testing. The design and conduct of international studies is an important part of discussions of national testing. A major challenge before all of us is in making better sense of the various testing and assessment systems at all levels -- local, state, national and international -- to make them efficient and to provide that results at any level may be related to results at other levels.

Mr. Chairman and Members of the Subcommittee, a week ago I had the opportunity to testify on the topic "Question of a National Test" before the Senate Subcommittee on Education, Arts and Humanities. In my statement before that subcommittee, which is attached to this statement, we recommended the answer to the question be a nationwide testing system. My comments today on international studies are made in the context of our proposal.

International comparisons in education have been and can continue to be based on assessments which sample student achievement. In international studies we are interested in knowing how entire systems or units of education perform rather than how each individual student performs. However, through international studies, we want to have a means by which achievement levels of students in any part of our country through individual examinations can be related to standards and results which are international in scope. International comparisons must be related to national standards which in turn are the guideposts for student performance in any part of the United States. Through the link of national standards related to international assessment, we can determine the relative performance of individuals in the United States to that of students in other countries. The key point here is the importance of having a process and entity for determining national standards for student testing.

I will not dwell on other features of the nation-wide testing system as described in the attached testimony, but I must emphasize the importance of constructing our national system so that there is an efficient connection with our national results to cross national or international comparative studies. This has not happened in the past. It will be a bold move for us to plan and implement a system which makes such relationships possible. But is is imperative that we design such an approach.

To expedite the participation of the United States in international studies, it is essential to plan for development of a nationwide testing system related to international studies as they may be used by the several states. Many states will probably participate in the 1993-94 Third International Mathematics and Science Study (TIMSS) just as the Canadian provinces will participate in addition to Canada as a whole. The states need plan carefully now for 1994 with respect to intentions in participating in NAEP, as well as in the international study. There should be a close relationship between NAEP and the international study and, therefore, we hope there will be early authorization to continue state participation in NAEP. I urge the committee to take action to continue authority for the states to use NAEP on a state-by-state basis in 1994 and beyond so those states which do choose to participate in NAEP and also hope to participate in TIMSS may design their overall state assessment plans accordingly.

In conclusion, Mr. Chairman, I note that studies done through IEA in the 1960s, 70s, and early 1980s, yielded some of the most important information for alerting the United States to the necessity for major reform and restructuring of education. Reports such as the Underachieving Curriculum: Assessing U.S. Mathematics from an International Perspective which came from the second international study of mathematics was very powerful in demonstrating the relatively poor performance of American students in mathematics. These studies have not just displayed achievement results, but they have presented important information about the time committed to instruction, the nature of the instruction, commitments for homework, commitments to standards expected of students which have sent powerful messages to consider here in the United States. We must increase the flow of information from such studies. We cannot be insular nor provincial in our approach to strengthening American education. One of the most important resources of information for reform here is in the examination of education in other countries.

Chairman, Members of the Subcommittee, thank you for the opportunity to appear before you this morning. I will be very pleased to respond to any questions about my remarks. Thank you.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

[illegible] Hall of the States 400 North Capitol Street, N.W. Washington, DC 20001-1511 • 202 393-8161 • FAX 202 393-122
Resource Center on Educational Equity 202 393-8159 State Education Assessment Center 202 624-77

"QUESTION OF A NATIONAL TEST" ANSWER: A NATIONWIDE TESTING SYSTEM

GORDON M. AMBACH, EXECUTIVE DIRECTOR
COUNCIL OF CHIEF STATE SCHOOL OFFICERS

MARCH 7, 1991

U.S. SENATE SUBCOMMITTEE ON EDUCATION, ARTS AND HUMANITIES

1. Chairman Pell, members of the Senate Subcommittee, and Staff of the Subcommittee, thank you for this opportunity to respond to your invitation to testify on "Question of a National Test." At the outset, Mr. Chairman, I recognize and commend you and your colleagues who have been advocates for national examinations long before serious consideration was being given to these issues for most of the nation. I note especially your responsibility for authoring the provision for national examinations in the Hawkins-Stafford Act of 1988 and commend you and members of the Subcommittee now for bringing the issues before the Senate in the manner of this hearing.

Testing is at the center of learning. The characteristics and manner of questioning or inquiry guides the form of education. How tests are constructed shapes what is thought and learned and, therefore, decisions about testing must be carefully made. That is especially true now because of the relation of testing to reform and improvement in American education and the location of education decision-making in the United States.

President HERBERT J. GROVER, Wisconsin Superintendent of Public Instruction • President Elect WERNER ROGERS, Georgia Superintendent of Schools • Vice President [illegible] • Directors [illegible] • [illegible] • JERRY L. EVANS, Idaho Superintendent of Public Instruction • BILL HONIG, California Superintendent of Public Instruction • [illegible] • GERALD N. TIROZZI, Connecticut Commissioner of Education • Executive Director GORDON M. AMBACH

(179) 25 [illegible]

Our American pattern of decentralization of authority for education in states and localities is being weighed off against important national goals and purposes for education. Who establishes standards and develops tests is, in many ways, as significant a question as the content of the test. The right combination of common standards with variations of tests which measure progress toward those standards is central to a national solution.

At the very time we search for successful reforms and improvements in education and debate the relative levels of decision-making at different points in the educational structure, we must also adjust to rapidly changing techniques of assessment. While we have a strong need to do things nationally -- to improve opportunities for students nationally and take steps to increase our national competence through education -- we must be certain to enable and encourage variation, experimentation and innovation that enables us to create and recreate ever better systems of testing and learning in the future while we are putting new standards and tests in place now.

2. Our response to a "Question of a National Test" is to recommend "A Nationwide Testing System" which has the following three key elements:

a) A procedure and national entity for setting national standards for student performance, subject by subject.

b) A system of both program assessments (through sampling of student performance) at national, state and local levels to determine program effectiveness and a system of individual examinations which might be nationwide, multi-state, state or local in administration which measure individual student progress on the national standards.

c) A reporting system which enables students, parents and responsible education officials anywhere in the United States to be able to relate any one student's achievement to the national standards and the performance of other students in the community, state, nation, and even the world.

3. The first element is to establish a procedure and an entity for setting national standards for the nationwide testing system. The United States currently has no national entity to establish national student performance standards. Such an entity, or Board, must be carefully designed and established through an Act of the United States Congress and the President. The Board should be comprised of distinguished persons appointed in equal numbers by the Congress and the President. Appointments to the Board should be based on a thorough nomination process which assures Board members will be well-qualified for their responsibility. The Board's responsibility should be to establish "frameworks" of student performance goals and objectives, or standards, upon which both program assessments and individual student examinations are based. The process of setting such frameworks must involve key education authorities at state and local levels.

Although it might seem establishment of such national frameworks is foreign to American education practice, the fact is that such a process is in place for the National Assessment of Educational Progress (NAEP). The process has already been used for the subjects of mathematics, reading and the sciences under the direction of the Council of Chief State School Officers as a part of the development of NAEP examinations. A similar process could be used by the new Board.

A second important responsibility of the Board should be review of various proposed program assessments and individual examinations in order to determine whether they effectively measure the standards established by the Board with reliability and validity. The Board would maintain quality control of various tests to assure rigorous measurement of both what students know and can do. The Board would exercise judgement on effectiveness of a proposed test for its intended purpose and the design for reporting results on the test.

At this time, Mr. Chairman, we are not presenting a specific bill for the creation of the Board, but we would be pleased to assist you and members of the Subcommittee and the Staff in the development of such a bill.

4. Program assessment components of the nationwide testing system might be at the national, state and/or local levels. NAEP provides the basis for the major program assessment component. For nearly a quarter century, NAEP has been providing periodic testing of samples of students across the nation. It provides overall trendlines for student achievement.

In 1988 Congress authorized the use of NAEP on a state by state basis. In 1990 the first use of NAEP on a state by state basis - mathematics at the eighth grade level -- was implemented. You have authorized state by state NAEP in mathematics at two grade levels and reading at one grade level in 1992. Authority for state by state NAEP, however, stops at that point.

We urge you take action as rapidly as possible to authorize the continuation of NAEP on both a national and state by state basis in five major subject areas -- mathematics, science, reading, writing and history/geography -- each to be tested every two years. We urge, furthermore, that you authorize voluntary participation in NAEP at a school district-wide level for those districts of sufficient size for appropriate NAEP sampling. We have a marked up bill which would accomplish the amendments recommended above.

Within the overall nationwide testing system, NAEP tests would be constructed to measure the subject objectives and standards established through the new Board frameworks described under point (3) above.

5. A nationwide individual examination system should be established within the nationwide testing system. This system could have several different forms of examinations. These might be used on a nationwide basis or by clusters of states or districts, or individual states or school districts. The determination of their use would be made by state or local education authorities. The examinations would be based on standards established by the Board under item (3) above. The types and forms of various examinations would differ, but, as noted above, to be part of the nationwide system they would have to be judged appropriate to measure the national standards by the Board. This system would enable creation of innovative forms of testing, including performance assessment, and enable a variety of approaches by states and localities in establishing individual examinations in an efficient and cost-effective manner. These examinations must be closely associated with the curriculum; used in a variety of patterns at different grade levels according to those points study at which subject mastery is completed; and be used as "high stakes" tests related to credit or credentials for individual students.

States and localities have a variety of existing individual examination systems, the most comprehensive now being the New York State Regents examinations. Different systems, such as that in New York, could be incorporated into the nationwide individual examination systems.

Of special note, Mr. Chairman, is the potential incorporation of the voluntary national examination authorized in the Hawkins-Stafford Act of 1988 into this system. This examination has never been implemented. It could be one part of the nationwide individual examination system, its content guided by the standards of a new national Board.

We recommend you take action to support the research and development needed to establish a nationwide individual examination system.

6. Common and consistent reporting of results from the different program assessments and individual examinations is of central importance in the nationwide testing system. There is now extensive testing at all grade levels for American students. The information from the tests, however, generally cannot be related so that results from one school or school district may be compared with another district within the state, or outside of the state, or outside of the nation. One purpose of the nationwide testing system is to create the means for relating results throughout the system without the necessity of requiring all students at all grade levels in all subjects to be tested on the same tests. Reporting systems must be established nationwide with information on student results related to other education indicators. This enables better understanding of the causes of student success or failure and helps results of testing lead to program and student improvement.

7. Testing systems carry significant costs in student instructional time, teacher time, and in the cost of creation, administration scoring and reporting of test results. Careful estimates must be made on the trade offs between program assessment sampling versus individual examinations and "every student testing." Careful estimates of cost must be made over this decade so that an efficient mix of program assessments and individual examinations is created and coordinated so as to limit costs locally, state by state, and Federal.

A nationwide testing system is essential. Information about the nation's education status is certainly as important as information about the nation's health, its agricultural condition, and the condition of labor and employment. In each of these other areas, the Federal government is now spending approximately six times as much for the collection of information about performance and system indicators as is true in education. The commitment to a nationwide testing system must be accompanied by a commitment to a Federal budget which makes certain the testing system is of as high quality as we expect student performance to be.

Mr. Chairman and Members of the Subcommittee, thank you for the opportunity to testify. I will be pleased to respond to any questions.

Chairman **Kildee**. Thank you very much, Gordon. I appreciate your testimony.

I'm going to start off with a general question, one that I have asked all the panels. I will shorten it up a bit.

Can we make testing, specifically national testing, helpful in the improvement of education?

Dr. **Cooperman**. I will give two examples, and they're both hopefully connected.

One is if we find out that by testing all the children—and I sure agree with Gordon, and I really agree with Governor Romer and Mr. Brock—that you really have got to know where you're going. You have got to know what it is you want to teach, and then have the standards.

Let's assume that most of the teachers in a particular State come from four, five, six or seven graduate schools of education. Let's further assume that the children aren't doing very well in math in fourth grade. In fact, they're doing kind of poorly. That would be a powerful message to go back and say you were not training your teachers correctly, because if the children aren't learning, maybe they're not being taught.

Let me give you a second practical situation. I asked the question why, in mathematics, are New Jersey's children and the children of the United States in elementary schools doing so poorly on every international comparison I could find? The answers were not difficult to find out. Most of the teachers had never taken a math course; most of the teachers—I hope this doesn't come out as sexist; I don't mean for it to—but most of the teachers in grades K-4 were women. They were told they were not good at math and so were not math majors.

The textbook manufacturers told me "we're not going to have a faster math pace because the teachers can't teach it." We taught in New Jersey 28 minutes of math—we conducted a study to find this out. We taught less math than anything else because the teachers didn't want to teach it.

Well, let's say that the results on a national achievement exam, or clusters of exams calibrated, that they show we're not doing well in math, specific and precise. Maybe what would happen is teachers would now take mathematics courses, that the school day would be extended, and more time would be allocated for math. There would be staff development or in-service training, and the curriculum would be more aggressive.

I tried something like that in New Jersey. I went all up and down the State and said we're not doing well. We're not doing well. People just turned away and said, "What are you getting so excited about?" Our graduate schools were mainly taught by Asians. The majority of the kids in mathematics graduate education were Asians. And yet there didn't seem to be much of a concern that we're not doing well.

Yes, I agree with Monty, that we have got to watch out, that we've got to have good tests and fair tests, and we can't abuse kids. We have got to do it right. I think we do have the ability to do that.

When the results come out, this will drive the ingenuity of people to figure out how to get things better. Right now there is a malaise. People will bait and hook you and try and get you talking about other things that we ought to do because they don't want the data.

I'll give you two examples. The data will go back to the teacher training and make that more powerful, and it will release a local ingenuity to figure out how can we do better once people are confronted and once they finally face the fact that we are not doing very well. We have got to face that fact first.

Chairman **Kildee**. Dr. Neill.

Dr. **Neill**. Yes, thank you.

Let me start by answering with a very simple example. If Maria can't multiply, why not? Well, the first one is let's assume she hasn't been taught. We don't need national exams for a teacher to know whether or not the teacher has taught Maria to multiply. So that's not helpful.

Let's say the teacher has taught Maria multiplication but it didn't catch. Why not? The current kinds of exams that are used in multiple choice tests provide no useful information on that topic.

The best thing to do, to start with, is to look at the kind of reasoning that Maria is doing mathematically. Here we need, of course, a teacher who is educated to be able to, one, know the kinds of questions to ask the student—and we don't do that a lot in our schools—and, two, to interpret that information into revising the kind of education material presented to Maria, the opportunities to Maria. It's not just the teacher talking but the different kinds of math experiences for Maria to go through in order to be able to understand finally how to multiply.

Testing is not a good mechanism for providing feedback for instructors. The kind of information that Mr. Cooperman was talking about can be provided by light sampling on the order that NAEP does. You do not need to test every child to find out, across the system, what kinds of things are going on.

The reason, in large part, to reform the kinds of assessment that we do is because performance-based kinds of assessments simultaneously involve instruction and assessment processes. But they are, as I said earlier, extremely difficult. We don't yet know how to turn those things into a nationally aggregated system.

So we then have a quandry. We go down the road of simplistic tests and that kills the education reform process, or we begin to do the more complex task, knowing full well that we don't have an outcome in sight. So the short answer is that most testing cannot help us in our education reform. Carefully done, some kinds of assessment can be used, provided that we have enough information surrounding it to make sense of the particular test information.

Chairman **Kildee**. Dr. Kelley?

Dr. **Kelley**. Yes. First I would like to point out that we don't use what's available with respect to current testing to help the teacher, as I mentioned earlier.

A second point is that all the problems mentioned that might be addressed by tests we know about, based on all the testing we do. We don't need another test to know that we have problems. Again, it's important to find out which of these kids don't know math, not what the average is. So I don't see another national test, particularly if we can define the problems carefully now—and I believe we can—of use in that dimension at all.

I do see some rationale for light sampling in terms of international comparisons. It's in a wholly different area. That is always the problem of educational finance. We have said that educating the 30 percent of kids who are not suitable for the modern work force, directly or indirectly, is going to cost money if we're going to maintain other values that we cherish in our society and which differentiate us from other societies largely. That money is hard to come by, and it's particularly hard to come by locally, where only 30 percent of families have kids in school. Any kinds of ammunition that can be used locally to provide the "Thomas Jefferson" effect, as I would call it—that is, education is good for democracy, and we're not doing well compared to other democracy type arguments with local citizens—is always good information to have on hand. But I see some value in international comparisons then for local financing of education, as long as it remains there.

Chairman **Kildee**. Yes, Dr. Hertz.

Dr. **Hertz**. I would just like to suggest that the further the testing is removed from this opportunity for it to feed immediately into the educational process, the less likely it is to be effective.

And then just one little side story. We may all want to watch Wisconsin with some interest over the next few years. Ten years ago in Wisconsin, a youngster did not have to take the ACT or the SAT or any other test in order to enter the university system, one nationally accepted as being absolutely excellent. Children didn't have to take a test to get in. There were no high school graduation requirements in Wisconsin. The local school district decided whatever it was that it liked for the kids to be doing. And yet, the children in that State have been doing beautifully, have been doing well, and there is an excellent work force there.

Now, in the past ten years we have changed all of that. We have built rather rigid standards into the expectations of the State, we have statewide graduation requirements, and we now are forcing children to take these tests in order to go into the university system. I think we all should look to Wisconsin with a great sense of history, and as historians we are going to want to find out whether we get better for all of that.

Chairman **Kildee**. Gordon?

Mr. **Ambach**. Mr. Chairman, I think your question takes us right back to the heart of the issue, and that is why it is that we need to be talking about nationwide assessment or testing systems rather than a single test.

I have to split the answer to your question in two parts—one, whether we're talking about national individual examinations of youngsters, or whether we're talking about national program assessments, which are the samplings.

As far as individual youngsters are concerned, any value of program assessments has to be indirect, if you will. Program assessments are designed to try to help us with overall policy, overall direction of what the program should be. The impact on a youngster is not going to be immediate but it is going to be in terms of how you use that information to inform a change in the program. So you're going to have that as an indirect. That's as very important function, though, at the national or State level by way of having that indirect impact on youngsters.

The key impact on youngsters, of course, is with individual examinations. But you have heard over and over this morning the difficulties in trying to pack it all into one single test. That is why our challenge, I believe, is to be figuring out the way that we can have systems through which there are many or several different kinds of individual examinations which get calibrated back to a common standard, that leaves you the flexibility of a direct effect for individual students in that particular examination, but gives you the advantage and, when the testing is done, there's a way in which you can connect the result to some kind of common standard, so that you know what's happened.

Greg Anrig was talking about what happens when the teacher closes the door, and that's correct. The concern is whether the results of what's happening when the teacher closes the door are the results we're looking for across that school, across that district, or in a broader scale. That's why you need to have some way of being able to judge whether that individual performance is up to or close to a particular standard.

Chairman **Kildee**. Dr. Cooperman, you had an addendum?

Dr. **Cooperman**. Thank you, sir.

What I have heard a lot of today, I guess, from the three gentlemen to my right, is frequently another test—my god, here comes another test—and therefore the test is painted to you that the whole idea of examination is bad and you should resist. Also an argument is put forth that we already have the test.

The tests that we have nationwide are basically three types: one is the Iowa Metropolitan or Stanfers. These tests are usually given to children in grades K-8. These are not secure tests, they're not called a high stakes test. The teachers frequently have them and coach the kids on them. And as John Cannell said on Lake Woebegone, all the kids are above average. He exposed that one. So first of all we have the Iowa-type test, which are basic skills in grade K-8.

The second thing we have is State testing, which is usually in high school, and is only basic skills. The third type of test is kids who go to college, ETS or ACT. We have no tests for all of our children that match and link—I'll use the word link—to our national goals. What's different now, and why we're talking about this now, thank God, in the history of our country, is we finally have national goals, not State goals but national goals—history, math, reading and writing.

I just think that a fragmented system which has three sets of rather inadequate tests can, through this discussion, be combined into a better system. Our way says, if you don't drive it back to each individual school—I agree with NAEP, I agree with what Gordon says, and I agree with what Governor Romer and former Secretary Brock said. You try and find out the answer to two basic questions: what is it you want children to learn, question number one, and question number two, how do you know that they've learned it.

My concern is, if we just do sampling, we will know how the country is doing. We will not know how that school is doing. The school, I'm afraid, in many areas will get off the hook and say the country might not be doing very well but I'm doing fine. I would ask you to please consider that.

Chairman **Kildee**. Mr. Petri.

Mr. **Petri**. This has been a fascinating panel, as well as the first one, and I'm very happy to have the opportunity to listen to all of you.

I have a lot of questions, but let me just ask one. If as a result of going to a national test on history, government, math, and so on and so forth American kids know more but know how to learn less well, will we have gained or lost?

In other words, is it all right, on average, to have American kids who don't know much in terms of specifics but know how to adapt and to learn better than, say, German kids or Japanese kids? That probably is the case on average, the way they perform when they're in armies and given a random task to do, to go out and do it and show some initiative. There is a lot of self-initiative in the American people that you don't see in other populations.

If we sacrifice some of that in order to do a better job in terms of imparting specific data at a particular point in time, have we gained or lost, or can we have both?

Dr. **Neill**. If I may go first, sir, we can have both. In fact, what tends to happen in this country, the CATS in metropolitan areas are extremely high stakes tests, and in many school districts their preparation occupies half of every school year. They are reducing learning to discreet little bits and kids don't learn the adaptable processes.

Let me give you an example. In California a study was done of the very best students in calculus who were high school seniors. What they found out in this study was that these students really do not know how to problem solve. They can apply a rule very accurately, when they know what rule to apply, but they don't know how to approach a problem in order to define the problem and then go about the process of solving it.

That's why I used my example earlier, that we need our students in history to think like historians. An historian will know the dates and the facts, but they will make meaning of the dates and the facts. If you just absorb dates and facts, you'll take one of those quick tests they give you three years after you leave school and, guess what; you don't remember anything. The Soviet kids don't remember anything, either. They got taught probably pretty much the same way.

If you use information and can think historically, you may not even remember all the facts—you probably won't. But the point is you will know the processes. Knowing information is not antithetical to the deep thinking. In fact, you have to think about something. But when we talk about learning, it's a matter of approaching children in the way humans actually learn, which is you adapt new information into the mental constructs you already have. It's a thinking process. You do drills and it makes sense to get better at something, but you put it in its proper place. We need assessment systems, not tests, that encourage thinking. If we have tests that don't do that, we won't get there.

Chairman **Kildee**. If the gentleman would yield just a moment, I taught school for ten years and a lot of times we test for the "what" and we don't go for the "why."

I used to draw on the blackboard the opening day of each class a triangle and ask them what the formula was to discover the area of that triangle. Most all the students would say one-half the base times the height. Then I would ask why is that the formula. It's amazing that most people didn't know why. When you complete it into a rectangle, you can empirically measure it, right? We have a formula when it's a rectangle. But I was amazed even then, that maybe in our testing we have to be careful that we don't just lead to the "what" and not to the "why."

Thank you for yielding. You may have some additional time.

Mr. **Petri**. Dr. Kelley.

Dr. **Kelley**. In response to the question, I think the extreme case you're talking about is the entering college exam in Japan. You look at where the Japanese send their engineering students when they actually want them to learn how to think, it's to American engineering schools. But more importantly, I don't think it's going to happen in the United States because, as contrasted to, say, when my parents went to school, if you have ever sat through a lot of classrooms, teaching of young children in particular is very much both and art and science now. These are people doing things that I don't know how to do, and I might be too old to learn how to do it.

Now, these people—there are a million of them—many of them have talked to each other in many professional associations now for years. It is wholly different from right after World War II. They have an idea about what these competencies should be, an they pay much more attention to that increasingly than they do to minimum competency tests and things like that. We are now dealing with people who actually should be here answering your questions, on the whole. In fact, it is somewhat amazing that I would be offering answers to your questions about this when people who actually think about this every day of their life, and think with their peers about that, would actually know much more about it, I think, and know much more about what is actually going on. I don't think they would allow it to happen.

Mr. **Petri**. I did have another question of my fellow "Badger"——

Chairman **Kildee**. Go ahead. I used part of your time.

Mr. **Petri**. I think the fact that here in Washington, although we're a smaller than average size State, there are more members on this panel from Wisconsin than any other State is just one sign of the value that our constituents place on education. I think we reflect that. We look in our State to education and value it very highly. It is what helps us to compete and do well. We're not perfect at all, but we try to work very hard at it, I think, on average in Wisconsin. We do devote a lot of resources at the State and local level to education, to put resources behind every student, regardless of the financial well-being of that community in our State.

The question I had was, I wonder if you would expand on something you mentioned almost in passing. You were thinking that maybe there was some hidden agenda with a national testing movement having to do with pushing choice or undermining the public school system, something like that. I wonder if you could spell out your fears or concerns, what it was you had in mind?

Dr. **Hertz**. It may merely be a reflection of an administrative instinct towards suspicion, Congressman. But as I was going through this last year, and I felt the enormous number of political forces—and I don't just mean State legislators on our Governor's commission, but from the unions and special interest groups of every kind, both of business and from the homes, the school people, all the way across the spectrum, impacting upon education, in a sense battering it from one side to the other. As I came away from that, I think I came away from it with a certain sense of suspicion about do all of these people, who seldom ever are in classrooms with children, are seldom ever in schools, are they all of good heart, or are they selling us something, or are they protecting something in terms of everything from jobs to whatever. That may be what you feel.

I don't know that I can point to anything specifically linking choice to the test question, but I have to say it's in my mind. I think that sense of suspicion negatively impacts—all of these things pounding away at public education I think does have a negative impact on us. I do suspect that in some cases the people that want a piece of the action—I don't even want to say that beat us down—but who want a piece of the action, want the schools to solve the social problems and all of the sort of things that are there, I'm not sure they're all of good will. I know many of them are, but I'm not sure that they all are. They all have the best interests of a broad cross section of children at heart.

Dr. **Neill**. Could I give a specific example and answer? In the Wall Street Journal the other day one of the arguments in favor of a national test advanced by Checker Finn was in order to have information, in order to allow school choice to proceed, and he also elsewhere has favored eliminating local school districts. So again, that fits into it.

I would just add that if we have a bad test, that will further discourage American education and further damage the possibility of enhancing public education.

Chairman **Kildee**. Mr. Hayes.

Mr. **Hayes**. Mr. Chairman, this has been a very, very interesting hearing here today. I could have several questions, but embellishing on what you said, Director Ambach, there is a limit that that part of the human anatomy which you sit on can endure.

[Laughter.]

We've been here now three hours and 45 minutes, so I just want to direct one question on something that bothered me a little bit. I don't know if any one of you can answer it.

We are facing a very critical situation in our educational system, what appears to be budding to the point where by the end of this century we're going to really face a shortage of African-American teachers. Do you have any suggestions as to how we can tackle this problem? There seems to be little or no desire for the kids in school to become teachers, particularly among African-Americans. Do you have any ideas on how we can tackle that problem?

Mr. **Ambach**. Congressman Hayes, Mr. Chairman, let me make a comment about one very important step that can and must be taken here at the Federal level.

Last year, toward the end of the session, your committee reported out the Excellence and Equity in Education Act. The House of Representatives passed it not only once but twice, with some revisions. It went over to the Senate and was killed by a few Republican Senators at the end on procedural grounds.

That particular Act carried, if my recollection is correct, more than $350 million for purposes of recruiting and preparing teachers for elementary and secondary schools and, in several respects, had a special emphasis on the necessity for the recruitment and preparation of Afro-American teachers, of Hispanic-American teachers. You have cited, quite properly, a very critical problem for the educational system.

I have said repeatedly since that event, which I believe was last October, that the first order of business for this session of the Congress should be to bring back that goals bill, bring back that Excellence and Equity Act, which the administration had agreed to finally, even after they had been here and telling you there wasn't any teacher shortage. They agreed to that. That bill ought to be put back in there in the fashion that it was before, perhaps with some alterations, and moved through as a significant act in order to address the issue that you're raising.

Dr. **Cooperman**. I have two ways. One is short term. You might check the alternate route to certification in New Jersey. I have a bias, like Gordon does, for the regents and so on in New York. In the past four years—and there's plenty of data on this—23 percent of the teachers in the alternate route to certification in New Jersey are minority. They have come from some of the more prestigious colleges in the United States.

I have met with the teachers, met with them for three straight years, and I asked them why do you want to teach in this State when you could have almost any job open to you? There were two responses. One, "I did not want to take 30 to 40 credits of teacher training in pedagogy." The second reason was, "I want to pay back and I want to make it better for some other kids."

Now, the average that are coming from the teacher training institutions across the United States is about 5 to 5½ percent. The program has been there for four years and I think there's enough data to show it does work. There is an appeal and they are getting people.

The other part of the answer, I agree with Gordon. You have got to have bold strokes. We were talking about testing here, and we may differ on testing. But that's not at the exclusion of other things that have to be done, to look at the research and to see what Larry Lazott and Brookover and other people have done with effective schools. Effective schools have been effective in urban areas, so why don't we replicate what makes an effective school? You have to follow Jim Comer to see what he did in New Haven and now what he's doing in Maryland. If it works there, it can work other places.

So I think what we've got to do is to see what has worked and replicate it, not not say it won't work here if it worked there. The other things is to put money into programs that have a good track record.

Chairman **Kildee**. Yes, Dr. Hertz.

Dr. **Hertz**. AASA has encouraged the idea of money going to minority students to be able to go to college. In our own community we wrote for some foundation money to do what we call a "celebration of teaching", Mr. Hayes. We take our children—and about 13 percent of our children are minority kids. We recognize the tremendous shortage in Wisconsin because we just don't have minority kids going through our university system. We know what the numbers are and they just aren't there. So we have to encourage the young people to go in. We have done that through this "celebration of teaching." We have decided to even move our efforts from the high school level, where we started them about four years ago, and move them back to middle school, because we feel we have to get them thinking about it earlier than when they get to high school.

So we started working with them at the middle school level, then at high school, and then we tell them, if you go off and pursue this thing of education in college, come back to us. We're selfish about the whole thing. But we think it's a motivation as well as a selfish instinct, you see. So it can work both ways. We say go off and do a good job, and then come back because there are going to be jobs here in our community. So I think they have to see a payoff at the end of whatever they're doing, and the payoff is a position.

Finally, I would just suggest that we are, through our process, trying to show them the economic advantages of teaching. I realize what has been said here, that those salaries have not gone up astronomically. But at the same time there is a good situation economically. There are a lot of fringe benefits and retirement things and we try to emphasize those things as well as being realistic about the salary.

Chairman **Kildee**. Dr. Kelley?

Dr. **Kelley**. About two weeks ago a survey came out on the educational and career aspirations of minority children. It wasn't a national but a regional survey. Quite surprisingly, there was a very large pool of people who were interested in going into public education. That was not true five years earlier. There is a pool of interest there.

The bill that did not get through the Senate last year actually, I think, needs to be looked at again, because nobody is going to go into debt like they would to become a lawyer to become a teacher, given the salaries that are paid. If you cost out what was in the bill last year, you still had to go into too much debt.

If I were reasonable and kind of knew what my career options were, I would become a lawyer and not a teacher if I didn't have the economic base behind me already. So I think you need to be aware of what those actual numbers are. You're spending money but not providing incentives, so you might get nothing out of it.

On the other end of it, it is pretty clear that one reason you don't have more minority candidates for professional positions all together is because, disproportionately, that 30 percent we're talking about come from urban areas and are minorities. That is a problem which I think we can identify and have to address, the education of all American citizens.

Again, dealing with that problem is not difficult. We know who the people are. We know at a very early age sometimes who the people are likely to be. We know what the problems are, and we obviously know what to do about it in many cases. We know it is not inexpensive in some cases. We know it's cost effective if they become employed, even in terms of the Federal tax exchequer. It has nothing to do with testing.

If we're talking about balancing, and we're saying all the parts of these 15, 19, 21 things we need to do to improve American education need to move in balance, I would like to remind you that, compared to other countries, we're out of balance. We test more relative to the other things that other countries do already.

Mr. **Hayes**. Thank you very much.

Chairman **Kildee**. Dr. Neill.

Dr. **Neill**. In addition to money and school reform, both of which are critical parts of this, there is one more. We currently have across the country tests that teachers in most States must take in order to become teachers. The most commonly used one is the NTE, the National Teacher Exam made by ETS. The pass rates for whites are typically in the 80–85 percent range; the pass rates for Latinos and African-Americans is typically well below 50 percent. That is, these students do not become teachers because they cannot pass the test.

Now, the claim of those tests is not that if you pass it you will be a good teacher, but that if you can't pass it, you don't have the ability to be a good teacher. That is a predictive validity claim. However, ETS has never done a predictive validity study to prove the point.

There is some evidence that exists, some from Montgomery or Prince George's County, MD over the last few years of where they had to hire teachers who hadn't yet passed the test. The ratings they got from their supervisors were as good as those given to those who did pass the test. There's a similar case in the military a few years ago; that is, the inability of tests to actually predict what we need to know.

The other thing that's happening with these tests is that there are many schools now, particularly black colleages in the South, that have reorganized their teacher programs. They have cut them drastically down to a quarter of the size they were before, and they reorganized the instruction solely geared to enabling these students to pass what is, by many educational experts, like a really bad test, the NTE.

I would say that the NTE is not anything itself that the Federal Government can do anything about. However, I would point out that pet food is more regulated by the Federal Government than the tests we give our children. I have nothing against my pet—I have a dog and I like her—but I do think we ought to think seriously about what we can do to improve the quality of the testing and that there ought to be something in place so that when a test has a predictive validity claim to it, you actually know it does the job.

Chairman **Kildee**. I think we will know this Nation has made a really strong commitment to education when in any given community the average teacher is making what the average attorney or average doctor is making. You know, you can judge a person's value system, a nation's value system, by how they spend their money. You go to the average community and the average teacher cannot live the life of the average lawyer or average doctor. I think that's a very important thing.

This has been an excellent panel. These last two days have been really helpful to the Congress and I'm sure to the Nation. I thank all of you. You're very, very busy people, people with great responsibilities, and the Congress and the Nation is grateful to you for your testimony here today.

Ordinarily I keep the record open for two weeks, but because I do want to get this record printed up, I will reduce that to one week for any additional submissions.

With that, the subcommittee will stand adjourned.

[Whereupon, at 2:05 p.m., the subcommittee adjourned.]

[Additional material submitted for the record follows.]

GOVERNMENT RELATIONS

STATEMENT
OF THE
NATIONAL EDUCATION ASSOCIATION

ON
NATIONAL STUDENT TESTING

BEFORE THE
SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND
VOCATIONAL EDUCATION
OF THE
COMMITTEE ON EDUCATION AND LABOR

U.S. HOUSE OF REPRESENTATION

MARCH 14, 1991

Mr. Chairman and Members of the Subcommittee:

The National Education Association which represents more than two million education employees in the nation's elementary, secondary, vocational, and postsecondary education institutions. We appreciate this opportunity to comment on the issue of a national student test.

NEA has long opposed a single, standardized national test for students. Already, students from kindergarten on are subject to a broad array of standardized tests -- required by local or state policy -- and of teacher-developed and administered tests. Adding another test to the mix would do nothing to advance any significant education goal, and, in fact, could undermine other important education objectives.

NEA believes that -- in the context of efforts to achieve the national education goals set by the President and the nation's governors -- it is appropriate to examine the whole area of assessment and accountability, and we agree that students, teachers, and schools should be held to high standards. We simply do not believe that the imposition of a national test is consistent with these aims.

Overview

The National Education Association recognizes the need for ongoing comprehensive evaluation of student progress. We believe that student learning must be assessed with measures directly linked to the lessons teachers teach and the materials teachers use.

Student assessment should not be equated with testing, and certainly assessment should not be equated with standardized testing. Appropriately, student assessment should involve teacher-developed and administered tests, formal and informal observation, and it should include assessment techniques that measure a broad range of skills -- many of which cannot be ascertained through a pencil and paper tests. Student projects, oral reports, and other assessment methods must be used to demonstrate mastery of subject matter. None of these techniques alone should be used as the single determinant of a student's future -- academic or otherwise.

Repeated studies demonstrate that standardized tests have inherent cultural and gender biases. All evaluation methods should be rigorously tested to eliminate such biases.

Opposition to national student testing should not be equated with opposition to holding students, teachers, and schools to high standards. NEA supports high standards and believe that local, state, and federal officials who share responsibility for educational policies must examine those standards on an ongoing basis to determine whether they are in accord with the current economic and political environment and with national and local needs.

Moreover, NEA recognizes that there is a place for standardized tests. But in order to assure that such tests are used appropriately, standardized tests must be developed, administered, and interpreted by trained personnel. Further, the results of those tests must be returned to the teacher and other appropriate staff in a timely manner in order to modify the instructional program.

NEA supports truth-in-testing as an important step for bringing about test reform, and we urge adoption of procedures that allow test takers to request a posttest review of all standardized test questions, scores, and rationale for the correct answers.

National Student Testing

The current drive to establish a national student test threatens to undermine the two chief strengths of the American system of public education: its diversity and its ability to give students a second chance.

Many proponents of a national test argue that a national test would merely set goals for a body of knowledge, and that local and state officials would have a high degree of discretion and autonomy to determine how best to achieve those goals. But the fact is that a single, national student test would lead inexorably to a national curriculum. If a national test is established -- with significant consequences for funding, personnel decisions, and curriculum -- then virtually every hour of every school day will be devoted to preparing students to do well on that test.

Not only would this tend to undermine the diversity of our public school system and the authority state and local officials now have, but it runs directly counter to the goal of better integrating higher order thinking skills -- such as problem-solving, analysis, and synthesis -- into the curriculum. Most nations which have a standardized national student test tend to use their educational systems to discourage diversity. While Americans investigate why it is that Japanese students, for example, tend to perform better on standardized tests, Japanese educators are doing research in this country to find ways to encourage greater diversity and creativity in their educational systems.

Moreover, nations that use a national test generally use those tests as a gatekeeper. A student's performance on tests at the secondary and postsecondary level can determine, once and for all, their educational and economic future. The United States places a high priority on individual choice and on giving individuals a second chance. In consultation with parents, teachers, and counselors, a secondary student in a public school in this country has a number of choices as to whether to pursue a vocational track, a basic educational track, or a college preparatory track. In Japan and most European nations, those decisions are determined by a student's performance on a test.

In the same way, Americans have a wide array of options in postsecondary education opportunities. In the U.S., an individual may choose to attend a college or not, may go back to school after a short or long hiatus, may pursue postsecondary vocational education or an academic education. As a percentage of the total population, more Americans attend postsecondary education institutions than any other nation. We make extraordinary efforts to provide that opportunity to individuals because we recognize that people's interests and level of commitment may change over time. A national student test would tend to diminish the ability of our society to provide those opportunities.

Improving Student Assessment

NEA strongly supports efforts to improve student assessment as a means of providing useful information about the effectiveness of educational practices and policies. Meaningful assessment methods should allow educators to diagnose student learning needs and select appropriate instructional strategies, as well as provide feedback to students and parents. They should enable school staffs to determine the effectiveness of curriculum and textbooks. And they should inform the community about how their schools are doing and what the schools' needs are.

If one can accept that these are the primary purposes of student assessment, it should be clear that traditional standardized tests have limited use for the changing needs of education. Rather than establishing a national test, Congress should support inservice education in evaluation techniques, research in effective assessment methods, and other programs that enable school faculties to improve the local assessment and, as importantly, to act on the conclusions of that assessment.

Conclusion

NEA believes that a national student test would only confirm what we already know -- that some students do better than others, some teachers do better than others, and some schools do better than others. NEA believes that a national test would reveal that schools perform best when they have sufficient resources to attract and retain quality staffs, have a broad academic program, and adequate equipment and materials. It would reveal that teachers do best when they have solid preparation programs, quality inservice education, and support from parents, the school itself, and ancillary agencies in the community. It would reveal that students do best when they come from families who support education and who have the resources to provide adequate nutrition, health care, and other human needs.

Achieving that difficult agenda -- assuring equity of educational opportunity through adequate resources and support from the community and the society at large -- is the only way to achieve the national goals in education established by the President and the nation's governors.

NEA believes that we have sufficient information to know what it will take to achieve those goals. All that remains is for us, as a nation, to provide the resources, programs, and policies to achieve them.

Thank you.

UNITED STATES CATHOLIC CONFERENCE
Department of Education
Office of the Secretary
3211 4th Street N.E. Washington DC 20017-1194 (202)541-3130 FAX (202)541-3322 TELEX 7400424

March 21, 1991

The Honorable Dale E. Kildee, Chairman
Subcommittee on Elementary, Secondary and Vocational Education
U.S. House of Representatives
320 Cannon House Office Building
Washington, DC 20515

Dear Representative Kildee:

As Secretary of Education for the United States Catholic Conference, I would like to take this opportunity to express the interest that our department has in the on-going deliberations of your subcommittee concerning the implementation of the National Goals for Education in general, but most specifically with regard to the question of a national test/assessment program.

The issue of standards and their assessment is a complex but important issue with many implications. There is a need to examine all of these possibilities carefully before taking any form of final action. One thing that I believe we want to avoid is a piece of legislation, along the lines of a single national test as opposed to a more flexible process of assessment. Such a single test might be viewed by some as a "quick fix" approach that could easily create more questions that it may hope to answer.

While I do not intend to speak for the total private school community, I believe that it is fair to say that we are greatly interested in these questions and have been trying to assess the implications on our students, programs and staffs. Representatives of the Council for American Private Education have been meeting with Archbishop Francis Schulte of New Orleans, who serves on the President's Advisory Committee on the National Goals, and have expressed some initial reactions relating to the implementation of the goals and specifically our concerns in the area of assessment. I believe that in the future, representatives of the diverse elements of the private school community may wish the opportunity to address these issues in detail with your subcommittee.

The private school community in general, and the Catholic school community in particular, plays a significant role in the educational well being of our nation. We have a strong interest in any proposal aimed at improving the education of our nation's young people. I believe that we can offer a valuable contribution to any debate on how to attain this admirable goal and therefore I believe that the private school community should be included wherever possible, in such deliberations and on committees which may be formed to address basic issues of interest to the entire educational community.

I hope you will place my comments in the record of the subcommittee hearings.

Please accept my thanks for your efforts on behalf of our school age young people.

Sincerely,

Sister Lourdes Sheehan, R.S.M
Secretary of Education

BEST COPY AVAILABLE

NATIONAL ALLIANCE OF BUSINESS

Chairman
John B. Carter
The Equitable
Financial Companies
President
William H. Kolberg
Board of Directors
Leo C. Beebe
K-Tron International, Inc.
Thomas D. Bell, Jr.
Burson-Marsteller
William D. Bonilla
Bonilla & Berlanga, Inc.
Edwin D. Campbell
Coalition of Essential Schools
James B. Campbell
MISSCO Corporation
James A. Cannavino
IBM Corporation
Bruce Carswell
GTE Corporation
John L. Clendenin
BellSouth Corporation
John J. Heldrich
Johnson & Johnson
Benjamin Hooks
NAACP
William H. Hudnut, III
City of Indianapolis
Elton Jolly
OIC of America, Inc.
[illegible] Kelley, Jr.
The Landmarks Group
Coretta Scott King
The Martin Luther King, Jr.
Center, Inc.
Jack Kirpinger
Wells Fargo, Inc.
Malcolm R. Lovell, Jr.
George Washington Univ.
Edwin A. Lupberger
Entergy Corporation
Jewell Jackson McCabe
National Coalition of 100
Black Women, Inc.
Ann McLaughlin
The Urban Institute
John D. Ong
The BFGoodrich Company
John E. Pepper
The Procter & Gamble Co.
Edward E. Phillips
The New England
Joseph A. Pichler
The Kroger Company
Ian M. Rolland
Lincoln National Corporation
Richard F. Schubert
American Red Cross
Barbara Shaw
NAPH
John P. Singleton
Security Pacific Corporation
John E. Sloan, Jr.
NFIB
Joshua I. Smith
Maxima Corporation
Paul G. Stern
NTI

March 13, 1991

Honorable Dale E. Kildee
Chairman
Subcommittee on Elementary, Secondary,
and Vocational Education
Committee on Education and Labor
U.S. House of Representatives
Washington, D.C. 20515

Dear Mr. Chairman:

The National Alliance of Business, in concert with the advisory board to its Center on Excellence in Education, has been working on the concepts of national student assessments and testing.

I would like to submit the attached written testimony for your review and as a formal statement for the record to be included with other testimony taken at your March 14th hearing on the subject of national student testing to evaluate educational progress.

I look forward to working with you and other subcommittee Members on these issues.

Sincerely,

W. H. Kolberg
William H. Kolberg
President

1201 New York Avenue, NW, Washington, DC 20005 202 289 2888 (FAX) 202 289 1303

BEST COPY AVAILABLE

NATIONAL ALLIANCE OF BUSINESS

Chairman
John B. Carter
The Equitable
Financial Companies

President
William H. Kolberg

Board of Directors
Leo C. Beebe
K-Tron International, Inc.

Thomas D. Bell, Jr.
Burson-Marsteller

William D. Bondis
[illegible]

Edwin D. Campbell
[illegible]

James B. Campbell
[illegible]

James A. Cannavino
IBM Corporation

Bruce Carswell
GTE Corporation

John L. Clendenin
BellSouth Corporation

John J. Heldrich
Johnson & Johnson

Benjamin Hooks
NAACP

William H. Hudnut, III
City of Indianapolis

Elton Jolly
OIC of America, Inc.

Blaine Kelley, Jr.
The Landmark Group

Coretta Scott King
The Martin Luther King, Jr.
Center, Inc.

Jack Kiepinger
[illegible]

Malcolm R. Lovell, Jr.
George Washington University

Edwin A. Lupberger
Entergy Corporation

Jewell Jackson McCabe
National Coalition of 100
Black Women, Inc.

Ann McLaughlin
The Urban Institute

John D. Ong
The BFGoodrich Company

John E. Pepper
The Procter & Gamble Co.

Edward E. Phillips
The New England

Joseph A. Pichler
The Kroger Company

Ian M. Rolland
Lincoln National Corporation

Richard F. Schubert
American Red Cross

Barbara Shaw
[illegible]

John P. Singleton
Security Pacific Corporation

John F. Sloan, Jr.
NFIB

Joshua I. Smith
Maxima Corporation

Paul G. Stern
NTI

STATEMENT FOR THE RECORD

BY

WILLIAM H. KOLBERG

ON BEHALF OF THE

NATIONAL ALLIANCE OF BUSINESS

BEFORE THE

SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION

COMMITTEE ON EDUCATION AND LABOR

U. S. HOUSE OF REPRESENTATIVES

"NATIONAL STUDENT TESTING TO EVALUATE EDUCATIONAL PROGRESS"

MARCH 14, 1991

The National Alliance of Business is an independent, business-led, non-profit corporation whose mission is to increase private sector training and job opportunities for economically disadvantaged and long-term unemployed individuals by building and strengthening public/private partnerships of business, government, labor, education, and community-based groups.

1201 New York Avenue, NW, Washington, DC 20005 202 289 2888 FAX 202 289 1303

BEST COPY AVAILABLE

STATEMENT FOR THE RECORD
BY WILLIAM H. KOLBERG
ON BEHALF OF THE
NATIONAL ALLIANCE OF BUSINESS
BEFORE THE
SUBCOMMITTEE ON ELEMENTARY, SECONDARY, AND VOCATIONAL EDUCATION
COMMITTEE ON EDUCATION AND LABOR
U.S. HOUSE OF REPRESENTATIVES

"NATIONAL STUDENT TESTING TO EVALUATE EDUCATIONAL PROGRESS"

MARCH 14, 1991

Mr. Chairman, I appreciate the opportunity to submit a statement for the record today to provide a business perspective on education reform and, in particular, about the need to assess our progress toward education excellence in this country.

While education remains largely a state and local responsibility in this country, we are now viewing education issues in a national context and as a national problem. This concern has led the National Alliance of Business to join in partnership with other national business organizations in the Business Coalition for Education Reform to work over the long haul with education and community leaders to help reverse declines in education quality and economic opportunity. (The Business Coalition for Education Reform includes: American Business Conference, Black Business Council, The Business Roundtable, Business-Higher Education Forum, Chamber of Commerce of the United States, Committee for Economic Development, The Conference Board, National Alliance of Business, National Association of Manufacturers, and the U.S. Hispanic Chamber of Commerce). American education quality is more widely, and more often, talked about among business leaders today than I would have imagined possible just five years ago.

The American education enterprise, which worked well for most of this century, has not kept pace with the increasing demands of modern society and our internationally competitive economy.

By way of illustration, the futurist Marvin Cetron has said that when the Class of the Year 2000 graduates from high school, the body of knowledge will have doubled four times since 1988. And these future graduates will have been exposed to more information during the Year 2000 than their grandparents consumed in a life time.

We are a society whose future relies heavily on the quality of our workforce. We must dramatically improve both how and what we teach for all our youth and equip them with the knowledge and skills for the world they will enter. It is equally clear, given future projections of fewer labor market entrants, that we can no longer afford to lose 30 percent of our youth who at some point drop out of school before high school completion.

Business has experienced, first hand, the results of lower American educational excellence. U.S. education performance has declined, when compared to other industrialized nations. This shows up in the workplace through lower productivity, lower wage growth, and a diminished competitive posture in the world market. This explains, in part, the compelling interest of business in education, although our interest as citizens in the quality of our society goes beyond the bottom line.

The National Commitment to Education

All of us just witnessed the dramatic events of war in the Persian Gulf. There are some valuable lessons from that experience about what this country can do when we anticipate and prepare for a potential crisis in a fast changing environment. We spent years on the development of new technology, on experimenting and refining new tactics, new strategies, and, in effect, modernizing every aspect of knowledge, training, and institutional structures for our national defense to meet new challenges in the changing world -- and the results were demonstrated. It shows that if we really put our minds and resources to something, we can be successful. We proved with the moon landing in the 1960s and with the recent Gulf War that when there is consensus in America, we get things done. Currently, we have neither national consensus, nor standards, nor assessment systems on what students need to know.

We in business now see our failure to educate our young people to world class standards as a major national economic problem which will be solved only if we apply the national will, new strategies, superior technology, and new institutional structures which can lead us to a comparable victory by the year 2000.

The industrialized nations who have become our key competitors around the world are those who long ago recognized the importance of education for their economic well-being. They each have adopted a national policy and practice for a systematic transition from school to work for all youth, and, with a high level of education skills in their workers, each has been able to organize work more efficiently with greater productivity by cultivating higher skills in front line workers. Their approaches to developing work opportunities through high skills and high expectations draws a sharp contrast to America's current approach.

National Standards and the Assessment of Progress

The National Alliance of Business and other organizations endorse the notion of establishing a framework for national student assessments. I need to emphasize the plural assessments -- not one test. We see this as a varied and cumulative system of assessments over time. And, it is a system based on a set of national standards. Our objective is for every student to leave school with a demonstrated ability to read, write, compute and perform at world class levels in general school subjects, and also to be able to learn, think, and work effectively both in groups and alone.

The concept of assessment which the Alliance supports would facilitate development of assessment systems that provide a variety of ways for students to demonstrate knowledge and skills. We do not want a single national test, but, rather, commonly held standards of performance and competence, demonstrated through a broadly understood system of diverse assessment techniques. Students would be assessed over a period of years. Cumulative assessment provides multiple opportunities for success rather than a single high-stakes moment of possible failure. It will enhance opportunities for the undereducated and under motivated to achieve high educational standards at their own pace, because the criteria for the assessments would not vary, regardless of the student's age. Although we advocate national standards of performance and a common assessment system, we also believe that the system should be administered locally, not nationally.

U.S. employers have no reliable gauge of what a new worker knows or doesn't know. Likewise, schools do not know what workforce skill demands are. Fortunately, we are finally seeing national and local efforts to bridge the knowledge gap between curricula learning and the workplace.

The National Education Goals Panel is developing recommendations on standards and assessment systems. The work of the Secretary (of Labor's) Commission on Achieving Necessary Skills (SCANS) will provide an important starting point by identifying the types and levels of skills needed for success in the competitive workforce. The National Center on Education and the Economy is developing a national assessment system which we think has real potential.

Once new assessment instruments are developed, based upon new goals and standards, they will take time to be widely disseminated, used effectively, and accepted. Even when implemented, substantial changes in student performance will not be quickly forthcoming. In fact, some of the new efforts may fail. Restructuring, by its very nature, rests on a willingness to experiment. Like businesses that restructure, efforts will not always succeed, but can point to more effective future practices.

Business leaders and educators are already combining efforts to define the standards of competence for entry into the workforce. Connecticut, New York, California, and Vermont, are developing new assessment tools and systems, ones that are so comprehensive in the information they capture that they can enhance student learning and can even form the basis for teacher and administrator assessment systems. They clarify learning goals, and more thoroughly measure the higher level skills that business and society require.

An assessment system should be viewed as something developmental that you learn from and refine as you do it. In fact, such systems should continually evolve, because the knowledge needed to succeed in life and work will change.

Changing the accountability system in education through assessment will also affect the management and administration of education, curriculum and instruction, and educational professionalism. The development of assessments cannot be separate from the professional development of educators, because it will have an impact on the nature of instruction and curricula content.

National Leadership and the Federal Role in Education Reform

We still lack many of the tools necessary to achieve broad-based and lasting success in our efforts to reform education. The establishment of national goals by the President and governors was a laudable first step toward defining national expectations for the education enterprise. These help to articulate where we want to go. Now we have to find the ways to build the national will to achieve these goals, and we need to be able to measure our progress along the way. Reaching the national goals must be a key national priority. I recognize the complexities involved, but we must articulate and apply a clear national strategy, have more cohesive leadership, and provide the necessary resources to succeed.

Currently, we have neither a national consensus nor standards, on what students need to know. We are told by teachers that they do not have a clear idea of what they should be teaching differently. We are told by parents that they do not know how their children are doing compared to other students, let alone among other industrialized societies. (In fact, parents are so confused that 75% of them believe there is a major education crisis, but 75% of them also think that the school their child attends is very good.) And students understand very little about what they will need to know to succeed in contemporary society.

First, national standards need to be established that set guideposts, a framework, and the expectations for what students should know. Standards should define what they need to know -- by age -- for being a citizen, for going to work, receiving additional skill training, or attending college.

We should have the ability to accumulate information about where individual students are in meeting these broad standards. That information should be gathered by school, by school district, by state, and nationally. Parents should be fully informed about how their children are doing.

The education standards we adopt should, of course, be competitive with other nations, but be designed to meet U.S. needs first. Our workers must be among the finest, if we expect our businesses to be as efficient as our international competitors, many of whom have national educational assessments.

The U.S., like Great Britain and other nations, need to address this problem directly. Britain has established a national curriculum that all students follow. It is standardized. Teachers have national criteria and assessments to measure student progress. Britain went at the problem directly by specifying curricula at the beginning, rather than having educational standards and assessments define the changes to curricula at a state and local level. What the British have done would be a revolutionary concept in this country. I use Great Britain to illustrate how strongly other nations are committed to the solution of their education problems.

We should be willing to challenge a few of the traditions in education by having a greater national approach, while still leaving most policy decisions to state and local levels. We've matured politically. We can talk now about national curriculum in this country, based on standards we establish. The important distinction is that this would not be a federal curriculum. I believe there is a role for federally developed criteria, standards, or syllabi against which state and local curricula can be developed.

It is my worry that we are not serious enough yet about making critical changes in education, and not committed yet to achieving our national education goals. For example, the fourth goal is to be first in the world in math and science education by the year 2000. For us to get there, it will require a national commitment comparable to the challenge of putting a man on the moon in the 1960s. I sense that we are still unwilling, as a nation, to think about the kinds of changes that would be necessary in policy, in structure, and in funding to achieve these goals. Without clear national and state leadership, educators are not likely to get behind the efforts with any sense of national mission.

The federal government, for its part, must be much more pro-active in its leadership to set the tone, set the agenda, and motivate action. There are several federal roles which can be played by the Administration and the Congress. The federal government should be undertaking a "moon landing" leadership role in developing standards for meeting our educational goals, so that there can be no doubt about what states need to do and where children rank in terms of what we have to achieve by the end of this century.

I firmly believe that there is a federal role in this effort as the lead partner in developing information, establishing standards, and devising assessments.

We are now at a juncture in our efforts to reform education where we recognize that we must act as a nation on a national challenge, and yet the needed actions must primarily be taken by 50 States and 15,000 local school districts, and, ultimately, in 83,000 school buildings. The federal government has historically been a relatively passive force in education and provides less than 10% of the funds for education. The challenge is to define a pro-active national leadership role for the federal government without displacing or supplanting the traditional State and local responsibilities for education.

In addition to our strongly held belief that the federal government must play a critical role in the development and establishment of educational standards and help to design a system of assessments, there are other specific activities appropriate to federal responsibilities in meeting our national education goals.

National Leadership through Research and Information Development. The federal government has a traditional, and accepted, role in supporting research upon which new knowledge and innovation in education are based. This role needs to be expanded in the 90's, as we search for new methods, develop standards and new assessment mechanisms, and to otherwise work toward meeting our ambitious goals. What is needed is a strategic approach to information development and the use of data in relation to national goals. We should carefully, but pro-actively, develop the national capacity to provide the knowledge and basic data from which the entire educational system can draw.

The single largest pot of research money that could be directed to these efforts is under the Education Department's Office of Educational Research and Improvement (OERI). What we should buy is a new strategy for the future education of our youth. The department should be an important catalyst for change, and should underwrite much of the developmental work that is necessary.

My sense is that the research agenda of this office has been to fund a broad range of small, discrete research projects which do not add up to a strategic plan or vision, does not provide information which the States and local schools need for their efforts to restructure education, and is not widely or systematically disseminated when completed.

We are missing an opportunity for the federal government to provide the leadership in research on education restructuring and national goals.

Early Childhood Development. This is a programmatic area in which federal leadership has had tremendous impact. It is the best example of filling a gap in the traditional systems of mandatory education. Early childhood development is an important new concept in education, as a strategy of prevention, of which the Head Start program is a part. The Committee for Economic Development (CED) has provided leadership on this issue, and has not only educated the business community about the importance of preschool education and health care, but also has argued convincingly for pursuing a strategy of prevention in public policy. On March 6, 1991, five Chief Executive Officers of major U.S. corporations testified before the House Budget Committee on full-funding of the Women, Infants, and Children (WIC) program.

We recognize how critical early childhood education can be. It has a direct impact on social skills, educational achievement, and self esteem. We at the Alliance see investments in early childhood programs, like Head Start, as an important weapon in the fight against the problems of school dropouts, drug abuse, crime, and teenage pregnancy and for that reason we are on record in support of full funding for Head Start. It has the potential, over the long term, of allowing us to redirect limited federal dollars that otherwise might have to be spent on "second chance" systems to repair the damage that could have been prevented. I would also argue that we are at a point where the costs could be shared with the States. We know that about 30 states have enacted various types of early childhood programs, 9 of which are designed to supplement Head Start.

Now, with the governors recognizing a national goal related to early childhood education, and with the states having primary responsibility for public education, perhaps we could move to a greater level of shared responsibility in early childhood development. There is precedent in virtually every other program of federal assistance to education. The closest comparison is the federal Chapter 1 program, which predominately covers poor children in the early years of elementary school, in which costs are shared with the states.

Build Linkages in the Broad Range of Federal Education Programs. We must take much more care to rationalize how individual programs are linked in a cohesive continuum of education development. Individual federal education programs must be thought of in relation to each other. For example, Head Start cannot be separated from services under Chapter 1 of the Elementary and Secondary Education Act or we risk losing the gains of one program during the application of another. Each program should build logically on the progress of the others. We still need to emphasize educational services to disadvantaged groups, but the policy must be to build on linking education programs. This also can have an effect in reducing the bureaucracies which have been established over time at the federal, state, and local levels.

Integrate Health and Human Service Programs with Education. Educators are often the first to identify health or other problems that are preventing children from learning. But, they are often helpless in addressing these problems. Federal and state legislation must put a premium on assuring that health and human service programs are readily available to children. This can be accomplished by insisting that such programs demonstrate how well they are tied into the schools and responding to the problems identified by school teachers, counselors, and administrators.

Establish a Better School to Work Transition. For several years now, various policy studies have focused on the failure of our society to provide school to work assistance to the majority of students who do not go on to college. We are not naive about how complicated this issue is, but it deserves urgent attention by this Subcommittee. I understand that several members of the full Committee have already expressed interest in working on such a proposal. In my view, it involves not only integrating opportunities for work experience with school to give relevance to classroom learning and to motivate students (as the Europeans do), but also to assist students with the skills for finding meaningful employment. America prepares only a tiny fraction of its non-college bound students for work. Other industrial nations have multi-year career educational programs that prepare students to operate at a professional level in the workplace. The enactment of the "Tech-Prep" or "Two-plus-Two" program in the vocational education reauthorization last year is an important step in this direction and may serve as a model for a more extensive system of occupational certification. The Alliance intends to develop more detailed ideas on the federal role in a school to work transition, and we will work closely with the Subcommittee during the this session as our work progresses.

Provide a Safety Net for Those Who Would Otherwise Fail. This is an important and traditional role of the federal government in education to assure equal opportunity and equity of services in education. We would like to see this role broadened in the way help is provided to school dropouts. There are a variety of existing programs in this area that need to be linked more carefully into a cohesive strategy. All students should be guaranteed the educational attention necessary to gain mastery of a standard set of educational skills by age 16, or as soon as possible thereafter.

Insist on Accountability. One critical lever that the federal government has over its investment in education is to carefully structure and insist on accountability. This means not only fiscal accountability, but also accountability for solid results. The Chapter 1 accountability standards are an example of what is needed. We are not prepared today to recommend specific methods to achieve accountability, but we do believe that rewards and consequences should be a part of education program legislation.

The Business Role

As for the business role, the business community must be included as a partner in this endeavor. Business can provide information on competency levels for jobs, so that learning experiences and curricula meet the needs of a dynamic society. Business can provide technical assistance in developing standards, assessment tools, and strategies. Business can work closely to help in training of teachers to apply these new assessment techniques. Business must also be an advocate for a comprehensive system of assessments and continue to press for systemic change.

Business leaders can be instrumental in keeping education high on the public agenda in their states and communities. They can be strong advocates for the transformation of the schools. They can help raise the sights of educators who, feeling powerless and frustrated, often lose any incentive to press forward. Joint efforts are necessary to address the spectrum of education issues in a coordinated and focused approach. Business leaders must work collaboratively and over the long term with educators as well as community leaders toward common goals.

In Conclusion

Mr. Chairman, this is a complex agenda for change. Despite what seem like insurmountable obstacles, a growing current of public opinion demands change in education. All Americans must play a part. We in business are preparing ourselves to play an important role in achieving significant change and improvement.

Our long term agenda, through the Alliance's Center for Excellence in Education and with our partners in the Business Coalition for Education Reform, is to find and implement more effective ways for business involvement.

This education reform effort requires strong national leadership from the federal government in setting the vision and the goals, fostering change, and in ensuring that all the stakeholders carry out their appropriate roles.

As America struggles to improve its education system, so that our society can be a strong and informed democracy and can compete in a global market place, it is essential that we are able to assess, in a meaningful way, the educational progress of our students. We are united in our determination to prepare our children to meet the challenges of the future. An accurate, and fair assessment system is an important tool toward that goal.

*****

Association of American Publishers, Inc.

Nicholas A. Veliotes
President

March 21, 1991

The Honorable Dale E. Kildee
Chairman
Subcommittee on Elementary, Secondary
  and Vocational Education
Committee on Education and Labor
320 Cannon House Office Building
Washington, DC 20515

Re: Comments of the Association of American Publishers on "National Testing" Issues

Dear Mr. Chairman:

During the March 14, 1991 hearing on "National Testing: Pros and Cons," the Subcommittee agreed to hold open the record until today in order to permit other interested parties to submit their views on the issues. The Association of American Publishers ("AAP") appreciates the opportunity to submit its views on national testing to this Subcommittee through this letter and attachments.

AAP is the principal national trade organization representing more than 235 commercial publishers of hardcover and paperback books; professional, technical and scientific journals; computer software; and educational and classroom materials, including tests and evaluation scoring materials. As such, members of AAP historically have been involved in meeting the needs of school districts and state education programs for various types of assessment instruments that have been used to assist in the measurement and evaluation of the nation's school children.

For the future, AAP stands ready to play an active role in the debate over national testing issues, so that commercial test publishers may contribute to a consensus solution for improving the nation's educational system. We recognize our responsibility to be as creative as necessary to help our society meet these new formidable and exciting challenges. In that respect, AAP's Testing Committee has already met with Governor Romer and intends to continue those productive discussions. We also will open a dialogue with members of the business community to discuss how commercial publishers may work with business leaders to better understand and respond to the perceived educational problems associated with the complex issues surrounding assessment.

NO SINGLE NATIONAL TEST

AAP strongly opposes any single test to be used nationwide. However, AAP supports the concept of a nationwide assessment program, provided it meets specific criteria that, from the experiences of commercial publishers, are required to ensure that all universally-administered assessments (as distinct from assessments of student samples which may be intended for use primarily to monitor school systems and/or trends) will be used to improve student instruction and school curriculum. In this way, AAP also believes that more accountability can be brought to the assessment system. At the same time, AAP firmly believes that test diversity and competition should continue to be integral to any reform in order to assure that the best quality products are available. Publishers want to assure this Subcommittee that the use of multiple tests can still provide uniform comparative information.

NATIONAL ASSESSMENT CRITERIA

1. There must be multiple testing instruments available. Different tests serve different purposes. Some tests are designed for diagnostic purposes, others for accountability reasons and still others may focus on monitoring progress. No one test, indeed no one type of test, can accomplish all of the diverse goals and objectives of this country's educational system. Nevertheless, as indicated above, multiple tests used in a coordinated fashion can still provide uniform comparative data.

A test should not be used for purposes for which it was not designed, nor should a test become the sole basis for determining how a student is doing. Consequently, AAP urges that the first important criterion for a national assessment program must be adherence to the concept that a single national test not be used for purposes of determining progress toward achieving national education goals and standards. Additionally, varieties of assessment formats, from normed-referenced, standardized tests to performance-based assessments, have a role in the mix. An explanation of the various types of tests and formats is included in the attached testimony of Burt Faldet, given before the Senate Subcommittee on Education, Arts and Humanities of the Labor and Human Resources Committee on March 7, 1991.

2. Each assessment must be demonstrated to be valid, reliable and bias-free. Scrutiny, criticism and debate about tests and the role they play in the lives of students and whether they are biased or exclusionary is not a new or novel development of the day. Those very issues have confronted the testing community, including developers, users, administrators and others, for years. As a

result, voluntary Standards for Educational and Psychological Testing were adopted by the American Psychological Association (along with the American Educational Research Association and the National Council on Measurement in Education). The Standards set forth the professionally-accepted constructs and procedures to be followed in developing psychological tests, determining their reliability and validity, implementing their use and describing research results.

APA's Standards have been widely cited with approval by Congress, state legislatures, courts and researchers. The most recent endorsement occurred during the House consideration of H.R. 1285, the Higher Education Technical Amendments of 1991, between Chairman Ford and Congressman George Miller (page H. 1812, March 19, 1991). Under these Standards, developers and publishers of any assessment instrument, regardless of whether it is multiple-choice or performance-based, must present research on scientific validity (whether the instrument predicts what it is designed to predict), reliability (whether its results are repeatable) and fairness to protected groups (whether there is bias against any identifiable group of test takers).

3. Any national testing program must encourage open competition among all test developers, including commercial publishers. Regardless of the types of assessments to be included in a national program, commercial test publishers have an extraordinary background and expertise in producing instruments that comport with the APA's Standards and with the equally applicable Code of Fair Testing Practices in Education. The Code is primarily focused upon professionally-developed tests used in and by educational institutions. It is the commitment to extensive research and development by commercial test publishers that makes them second to none. Because these are private sector companies, they invest significant monies each year in conducting scientific research.

Similarly, this Subcommittee should be aware that AAP members have been involved in offering performanced-based assessments (e.g., writing and listening assessments) for years. Some publishers are also offering portfolio assessments as supplements to current test batteries. Accordingly, commercial publishers have already demonstrated the initiative and responsiveness to the demands for performanced-based assessments. Any national program should seek to continue an open competition in the development and administration of assessment instruments to ensure that the best products are available to meet emerging educational needs.

4. A national assessment system must supplement, not supplant, existing state and local assessment. AAP feels strongly that control over the attainment of national goals and selection of

assessments must remain with the state and local authorities. These entities are in the best position to make the necessary refinements to any national assessment program that will ensure that it is tailored to the needs of the local community which is being served. As several witnesses have previously testified before this Subcommittee, this flexibility is important to preserving the best in the American educational tradition. A national assessment program should not replace state and local testing programs, which may serve other purposes, particularly diagnostic.

5. State and local entities must integrate the national assessment system with on-going assessment. Although local control must reside with the state and local education agencies, they should have in place an explicit and comprehensive program to integrate and coordinate a national assessment system with on-going assessments. AAP submits that such an integration approach will provide a mechanism to ensure that all universally-administered assessments will be linked to the improvement of instruction.

6. The assessments must be administered and the results interpreted in accordance with the APA's Standards and the Code. Again, reliance on the adopted standards already in place must be assured, so that all aspects of the national assessment program can function together. Accordingly, it is important that test administration and test interpretation equally meet the high standards currently set forth by the APA's Standards and the Code. It makes no sense for the development of various assessments to be predicated on these standards and not have that action followed up by requiring those who administer the tests and who evaluate the test results to conform as well.

7. Much more emphasis must be given to assessment training. Further to the notion that test administration and interpretation must be as rigorous as the development of the national assessment program, AAP recommends that state and local education agencies should have in place a program to train teachers and administrators on the use and interpretation of assessments, including but not limited to the national assessment system. Research studies have shown that teachers do not make full use of test results, simply because they have not received adequate training on these principles. Out of this point also derives the added benefit of improved accountability, since better informed teachers and school administrators will be better able to report the results of assessments to pupils and parents. In fact, the general public may also benefit from the greater awareness and understanding of the role of testing in our society.

KEY ISSUES FOR A CONSENSUS

Before any national assessment system is implemented, however, the AAP urges this Subcommittee to pay particular attention to the need to reach a consensus on several key items that will be crucial to the success of the national program. Such a consensus must involve the participation of members of all affected communities of interest, including experts in assessment and measurement. For its part, AAP is earnestly desirous of contributing manpower and resources to such an effort and we would welcome the opportunity to discuss ways for accomplishing this avenue for national consensus with this Subcommittee or any of the groups that appeared at the hearing.

The following issues are those upon which agreement should be reached because they appear to AAP to be relevant to the development of a national assessment program that can and will achieve the goals and objectives advocated by the hearing witnesses:

1. The content, skills, objectives, and sequence to be taught and assessed, as well as on the interpretation of test results, e.g., attainment or achievement levels, pass/fail levels, college bound vs. technical, etc.

2. The purposes of a national assessment program, e.g., system accountability; monitoring progress in the attainment of national education goals; evaluating the results of the effectiveness of instruction; evaluating and improving curriculum; diagnosing individual strengths and weaknesses; or certifying the levels of achievement and ability upon ending secondary education. Subsidiary to what purposes are served, AAP feels compelled to raise the point that "high stakes" accountability testing may provide great leverage for changing the educational system, but it must be approached cautiously so that there are not distortions or unintended consequences and so that it will produce improvements in instruction and learning.

3. The voluntary or compulsory nature of participation in a national assessment program.

4. Who pays for the development, administration, scoring, and other costs of the program.

5. The scale of assessment, e.g., testing all students or a sample; testing annually or at intervals; testing all grades or certain grades; testing all students or excluding or providing separate tests or scores for some (e.g., non-public schools; Chapter 1; learning-disabled; ESL; etc.).

6. The frequency of administration and of revisions of the assessment to assure comparability of results over time, while encouraging a dynamic and diverse assessment program.

7. The security or accessibility of the assessments, i.e., will test items be secure or released to the public; will test items be protected by copyright; will test items be released to be imbedded in other assessment instruments.

8. The equating or scaling of different assessments, particularly assessments using different formats (e.g., multiple-choice or performance assessments), in accordance with the Standards. On this point, AAP notes the difficulty of the technical task of equating or calibrating different assessment instruments to an "anchor" test is not to be underestimated.

9. The choice of format of the assessments, (multiple-choice, performance, portfolio, or combination), is a function of a variety of factors, including cost, purpose of the test, timing, etc. Each format has advantages and each has disadvantages. It is not accurate to state that a performance or "authentic" assessment is the "best" assessment format without qualification.

On this last point, AAP is in extreme disagreement with those who advocate the need to completely remove the use of standardized, normed-referenced tests. That is akin to "throwing out the baby with the bath water." The use of multiple-choice tests has produced useful information on an objective, valid, non-discriminatory basis. This Subcommittee should not, any more than some educators and researchers, "shoot the messenger" when the message may not be to their liking.

DEVELOPMENT OF NATIONAL STANDARDS

Finally, AAP assumes that this Subcommittee recognizes that the task of developing the instruments for use in a national assessment program is secondary to the task of establishing national education standards. Here again, it is the challenge for every group involved in the educational system to contribute to a consensus-building effort as to what every child should know in any particular subject area in any particular grade level. AAP hopes that this Subcommittee will assure that this effort does not become the classic case of "putting the cart before the horse." It is premature to develop an assessment system until a consensus is reached about what is being assessed.

AAP believes that the participation of commercial publishers in this important task can be utilized at critical points in the formulation, development and delivery of appropriate assessment instruments and to ensure that instructional materials are developed and available. Even though this effort promises to be lengthy, perhaps even on the order of the ten-year program development outlined by Dr. Lauren Resnick, commercial publishers are committed to assuring that the national assessment program will be administered, and the results interpreted and used, in a standardized, unbiased way.

CONCLUSION

For all the above reasons, AAP respectfully suggests that the Subcommittee's attention should focus first on the need for a consensus of education standards and then, secondarily, on the need for a national assessment program, rather than on a single national test. AAP members, who currently comprise the largest segment of the assessment delivery system in place in this country, are willing and able to respond to the emerging needs of a changing view of what multiple assessment instruments, and possibly their delivery techniques, will be demanded by the year 2000, just as they have done over the last quarter of a century, when numerous "new" assessments and vogue education reforms have required similar commitments of resources.

If further hearings are scheduled, we would welcome the chance to report to this Subcommittee on the progress of these efforts. In addition, we would be pleased to answer any questions this Subcommittee may have based on these comments.

Respectfully submitted,

ASSOCIATION OF AMERICAN PUBLISHERS

Nicholas A. Veliotes/rt

Nicholas A. Veliotes
President

Attachments

Association of American Publishers, Inc.

aap

1718 Connecticut Avenue, N.W. #700
Washington, D.C. 20009-1148
Telephone 202 232-3335
FAX 202 745-0694

STATEMENT BY BURTON W. FALDET
ON BEHALF OF THE ASSOCIATION OF AMERICAN PUBLISHERS
BEFORE THE SUBCOMMITTEE ON EDUCATION, ARTS AND HUMANITIES
COMMITTEE ON LABOR AND HUMAN RESOURCES
MARCH 7, 1991

Mr. Chairman and members of the Committee, by name is Burt Faldet. I appreciate this opportunity to appear before you today on behalf of the Association of American Publishers. The Association of American Publishers ("AAP") is the principal trade organization representing more than 235 member firms that publish hardcover and paperback books; professional, technical, and scientific journals; computer software; and classroom and educational materials, including tests and evaluation and scoring materials.

I am President of Test Consultants, Ltd, which provides evaluation, design, and implementation strategies to education and business. Our clients have included commercial test publishers, the American Institutes for Research, IBM, as well as individual school districts. From 1965 to 1987, I was with Science Research Associates, a commercial test publisher, where I was involved in a variety of positions in the development, publication, and use of standardized tests for schools and industry. I also have taught undergraduate courses in Measurement and Evaluation and secondary school science, and served as a School Psychologist and Director of Pupil Personnel Services.

There are several points that I would like to discuss today, from the perspective of the publishers of such tests, about the development and use of standardized tests in elementary and secondary schools -- whether for the purpose of evaluating progress toward national education goals or to an individual pupil's development. My statement does not address higher education, employment, or military testing.

The first, and what I hope will be the most important message I leave with you today, is that developers and publishers of standardized tests should be seen as part of the solution for improving the quality of educational instruction, not as part of the problem.

The second message is that test diversity and competition should be encouraged to assure improved education and improved assessment instruments. Different objectives are served by different kinds of tests -- no one test can accomplish all of the diverse objectives of our diverse educational system. It is a serious mistake to try to make tests do what they are not designed to accomplish, or to use tests as the sole means for assessment in most situations.

Finally, I want to assure the Committee that test publishers -- working with the educational community -- are expanding and improving their testing products to meet continually emerging educational demands. Test publishers are ready and able to develop and administer whatever assessment instruments are desired, including assessments to measure progress toward attainment of national education goals.

WHY TEST?

Measurement can be relatively exact -- but a number has no meaning until someone makes a judgment about it. That is the difference between measurement and evaluation. There are many ways to determine health; a number on a thermometer is one indicator, but it takes someone to exercise judgment as to the significance of the temperature shown, and to take the appropriate action as indicated by the reading on the thermometer. It would be imprudent, however, to rely entirely on temperature to make a diagnosis of the patient.

Why educational testing? Testing is of value to the student. It serves to provide some information that can be used by educators and parents to identify and respond to the instructional needs of individual pupils and to improve instruction of individual pupils. Testing is a means to assess progress toward specific educational objectives, as evidenced by what pupils can do in terms of skills exhibited.

Testing also serves broader, institutional goals. It assists in assessment of long-range effects of changes in the educational program, enabling comparison of (1) performance over time and to changes in the instructional program or to changes in population characteristics and (2) performance across different subject areas, such as mathematics and reading, to determine strengths and weaknesses, needs for program modification, or changes of emphasis. Testing is one means to evaluate performance for accountability purposes.

The methods of evaluating whether children are learning what is being taught have changed over the years, just as many techniques and objectives of teaching have changed. For example, standardized achievement tests and numerous other types of tests have supplemented teacher-made tests administered on a class-by-class basis; performance and portfolio tests are serving to complement and supplement multiple-choice formats.

LIMITS TO TESTING

It must be emphasized, however, that there are limits to testing. When testing is used in "high-stakes" situations and results are used as a simple "pass/fail" barrier to students, or to reward or punish teachers and administrators, when the pressure becomes so intense that there is "teaching the test" rather than teaching the skills and concepts that are being evaluated, when test scores become the sole criteria for evaluating student performance or potential or the effectiveness of instruction, then testing has gotten out of hand and is being misused and abused.

Tests are a necessary but not sufficient means to assess achievement and growth in skills and abilities. What may be tested is not, and cannot be, inclusive of all of the desired outcomes of instruction.

Tests may be used as a partial basis for evaluation. Tests are concerned only with certain basic skills and abilities and are not intended to measure total achievement in any given subject or grade; they are not inclusive of all the desired outcomes of education. Standardized tests are concerned with only those areas of instruction that are amenable to objective measurement.

It should also be recognized that local performance is conditioned by many influences. The instructional effectiveness of the teaching staff is only one of these factors. Among other factors are the pupils' school and home environment, their past educational history, and the quality and adequacy of the instructional materials with which the staff has to work.

As stated in the Manual for School Administrators for one standardized test,

> At all times, the tests must be considered a means to an end and not ends in themselves. These tests have their principal value in drawing attention of the teaching staff and the pupil to those specific aspects of the pupil's development most in need of individual attention; in facilitating remedial and individualized

instruction; in identifying those aspects of the whole program of instruction most in need of increased emphasis and attention; and in providing the basis for more adequate educational guidance of the individual pupil. If properly used, the results should motivate both teachers and pupils to increased, better-directed efforts in both teaching and learning.

When intelligently used in combination with other important types of information, the results obtained from these tests should prove very valuable in the appraisal of the total program of instruction. Unless they are used in conjunction with other information, however, they may do serious injustice to many teachers and to many well conceived instructional programs.

KINDS OF TESTS

Different tests have been developed to meet a variety of purposes. Some tests are subjective, both as to the matter tested and the interpretation of the results. A standardized test is an objective test that uses the same standards to measure and score student performance across the country; everyone takes the same test according to the same rules.

A normed-reference test (NRT) is a standardized test used to compare students' performance in terms of a carefully selected, nationally representative group, or norm, on the same test; performance is based on total test or subtest scores. (In contrast, for some tests, such as the SATs and ACT, the norm is based on the others taking the test, rather than to a standardized national norm.)

A criterion-referenced test (CRT) differs from a normed-reference test primarily in how test scores are interpreted and used. A criterion-referenced test is used to evaluate and report performance in terms of specific instructional objectives or skills, stated in measurable terms.

These labels are not mutually exclusive. Many criterion-referenced tests are normed, and many norm-referenced tests may be subject to criterion-referenced, content-based interpretations.

Teacher-made tests generally are intended to provide information about individual student's performance on specific, classroom-oriented, curricula or specific needs for information about students. These tests are frequently supplemented by textbook tests, which are developed by textbook publishers and may appear in textbooks or be provided to teachers as supplementary instructional materials. Both of these tests are associated frequently with grades on report cards and help measure a student's progress in class, as well as facilitate individualized instruction.

Tests can also be in a variety of formats. Multiple-choice tests offer the advantages of objectivity and uniformity of scoring, ease of administration and scoring, and low cost. There are disadvantages to such tests, particularly if they are utilized as the exclusive method of assessment. "Performance-based tests," "authentic assessments," or "alternative assessments" generally are open-ended tests that are not multiple-choice. They include essays, writing samples and portfolios of work, practicums, or oral or visual demonstrations. They generally are more expensive, labor-intensive, and require more training and preparation to administer and evaluate -- factors which also can make them affirmative educational tools. The same concerns for fairness, validity and reliability, standardization if used for comparisons, and abuse if used in high-stakes situations that are raised with multiple-choice tests are applicable to performance tests.

Performance testing and standardized testing are not mutually exclusive. It is important to point out that for several years writing and listening assessments -- performance tests -- have been offered by test developers as part of their standardized test batteries. Publishers are now offering performance and portfolio tests to supplement their current test batteries.

What are the particular advantages of a norm-referenced, standardized test? It ensures reliability and validity in data collection, analysis, and interpretation. It enables evaluation of student achievement in various grades and subjects for the purpose of aggregating and reporting achievement gains in terms of a common reporting scale (e.g., normal curve equivalent or grade equivalent), with nationally representative norms. It provides an objective, rather than a subjective, assessment. The efficiencies are greater and the costs of administration are far less than for a performance assessment (by perhaps, a factor of 10).

Norm-referenced, standardized tests also enable identification of problems in specific skill or subject area deficiencies for teacher attention and remediation. This may be particularly important in the early grades.

Norm-referenced, standardized tests use the same or parallel test items for all students, which makes scores for all students comparable; use of one level per grade facilitates criterion-referenced interpretation of results for classes, buildings, and systems. Individual scores can be related to comparable national norms. One skill can be compared to another on a pupil, class, building, or system basis.

A classroom may have such a wide range of skills that no simple test can be equally suited to the entire range of achievement; NRTs for different levels of achievement can be administered so that each pupil takes the level that corresponds most closely to the individual instructional objectives and levels of skill development.

ROLE OF THE TEST DEVELOPER AND PUBLISHER

Test developers adhere to strict standards, as developed by the American Psychological Association, the American Educational Research Association, and the National Council for Measurement in Education in the Code of Fair Testing Practices in Education, a copy of which is submitted for inclusion in the record. Demonstration of reliability and validity also must be provided to test users, showing that the test meets its intended purpose and its appropriateness for groups of different racial, ethnic, or linguistic backgrounds who are likely to be tested. Several books give in-depth, candid reviews of available tests, include the Mental Measurement Yearbook, published by the Buros Institute of Mental Measurements, while guides and evaluations are published by the ERIC Clearinghouse on Tests, Measurement, and Evaluation and by other organizations.

Standardized tests generally are professionally developed tests distributed by commercial test publishers; development may be by the publisher, educators or other non-profit organizations (under royalty or other forms of compensation), or by governmental entities alone or in cooperation with publishers (such as under the National Science Foundation's "Publisher Initiative").

The role of the commercial test publisher in test development is very extensive. Based on information from a variety of sources, including the educational community, the test publisher determines if there is a need for a test and whether it will be financially viable. If the answers are in the affirmative, a decision is made as to the type of test to be developed, i.e., a norm-referenced or criterion-referenced test, or a combination of the two. In addition to the type of test, the format (e.g., multiple-choice, true-false, performance assessment) also must be determined. Publishers also respond to test requirements of state and local education agencies. Test publishers will also respond to the demands and requirements for tests of national education goals.

Extensive research is required for "building" a new test or revising an existing test. Test items are written by educators and professional test item writers. They are selected after extensive research on educational objectives; curriculum; goals, objectives, and standards; textbook and instructional material content; and what is to be measured and how. In general, tests follow and reflect curriculum. Error-free items must be developed that will withstand the scrutiny of hundreds of thousands of teachers and students over a long-period of time. Vocabulary and readability levels must be appropriate for the students to be tested. Items must also be free from ethnic, gender, or cultural bias.

At least one tryout to obtain data for standard item analysis and summary test statistics is needed. This data is used to select items with desirable characteristics. Typically, an experimental edition will contain at least twice the number of items required for the final test, to enable the publisher to reject undesirable items and still retain a sufficient number of items for a final test of suitable length. Items for a norm-referenced test will be rejected if too many examinees select the answer. In a criterion-referenced test students are classified in terms of mastery/non-mastery, so items will be selected that will have a large number of correctly-selected answers.

Experimental test items are reviewed by educators and curriculum specialists and are then field tested with large numbers of student to check their responses. The comments of the reviewers and the data generated by the field test are used to select the items for the final edition of the test.

In the case of norm-referenced tests, the final, or standardized, version of the test is administered to carefully selected groups of students whose characteristics are similar to those of students throughout the nation. The information obtained is then aggregated into norms so that individuals tested in the future may be compared to the original national sample. This is the process of standardization, and the normative information obtained from the process is crucial to educators, parents, and students. Without it, there would be no way of knowing how a single score on a specific test compared to the scores of other students in the nation.

Publishers develop guidance materials to assure that the final test is administered in accordance with the standardization, and to provide instruction on how the test is to be interpreted. Information is also developed and provided on the technical characteristics of the test to support its reliability and validity. This is done regardless of the format.

Scores can be reported and evaluated in a multitude of ways, for different uses. Rather than trying to describe scoring and interpretation in my testimony, I am submitting for the record an excerpt from Understanding Achievement Tests: A Guide for School Administrators, published by the ERIC Clearinghouse on Tests, Measurement, and Evaluation, on "What Types of Test Scores Are There."

Much controversy has been generated recently over norm-referenced testing. To address these concerns, I am attaching to this statement several articles from commercial test publishers that were included in the Summer 1988 Educational Measurement: Issues and Practice that provide an extensive review of these issues.

WHAT SKILLS ARE TESTED?

Higher order skills, not just basic skills, can be measured, even in a multiple-choice format, in a standardized test (remembering that it was only a very few years ago that publishers had to respond to demands for assessment instruments for the "back to basics" movement). We recognize, too, that there are more direct ways of measuring higher order skills and such instruments have been available from test publishers for some time.

As previously stated, the multiple-choice format used in assessment instruments has some attractive features. It is an efficient and effective way of measuring many educational objectives. While we recognize that it has limitations as well, it is important to recognize that most measures, including criterion-referenced and performance tests, are samples of behavior from which inferences can be drawn. For example, a multiple-choice mathematics test, which includes five exercises in addition of two-digit numbers with carrying, is a sample of all the possible two-digit numbers that we want a student to be able to add. For efficiency, we chose five exercises, and based on the student's performance on those, we infer what the student could do if presented with many more. Similarly, we may present a situation with several complex problem-solving exercises in a multiple-choice format. Based on performance, we can make some inferences about the student's performance in some of the higher order skills in the mathematics area.

Similarly, we can infer some important aspects of performance in writing from items commonly presented in multiple-choice language arts tests.

Neither the problem-solving nor language arts tests are substitutes for direct observation of student performance over time and in different situations in solving problems and in producing written material.

Reiterating a constant theme of this statement, that tests need not be mutually exclusive, I again want to point out that publishers of standardized tests currently also offer a variety of performance tests (including listening skills and writing) as well as portfolio programs, in addition to multiple choice tests.

Whether multiple-choice or performance tests, the keywords for the future, as they are today, are validity and reliability. Publishers cannot and should not market a test unless it has been demonstrated to be valid and reliable. This requires time and money, extensive research and development, testing and reworking to assure that the test works.

RECOMMENDATIONS AND SUGGESTIONS FOR FEDERAL POLICY

On behalf of the publishers of standardized tests, I welcome this opportunity to meet with the Committee and discuss standardized tests and our role in the educational process. As I said at the beginning of my statement, publishers want to be part of the solution, not part of the problem. Publishers are not simply printers, bookbinders, and marketers. They are an integral part of the educational system, providing an essential delivery system as well as taking the initiative for and bearing the risk of developing new and innovative materials. Just as Congress would not think of addressing the future of the automobile without consulting with automobile manufacturers, publishers should continue to be consulted and included in your continued deliberations over the quality of education and the development and assessment of national education goals.

What recommendations do we have for Congress? The first is that you continue to hold hearings such as this on education issues, particularly testing, as a prelude to any possible future action.

Second, Congress should continue to assure diversity of testing. No single test, no single curriculum, no single textbook, can or should meet our nation's diverse educational needs. Competition among test developers, including a vigorous private sector, should be encouraged.

Third, publishers have a role in making whatever testing program that may be adopted by a school or for a national education goals program work. They provide an economical and efficient delivery system for assessments. Publishers have traditionally served as an important bridge between sound theory and sound practice. Indeed, they have been the vehicle for getting local school acceptance of new concepts and the resulting products, and for enhancing and modifying those products as needed. They have been the primary link between those who create and those who must implement. We do not see a change in this role, nor do we believe that a change is desirable. For this reason, it is important to involve the publishers early in the conceptualization of products resulting from sound research.

One of the crucial concerns is the proper interpretation of test results. Our fourth recommendation is that Congress consider funding for targeted, in-service training to teachers and administrators in interpreting test results to enable them to use tests better to improve instruction, and to convey information to students, parents, and the public.

Fifth, state and local education agencies might be required to develop a comprehensive assessment plan, which would identify instructional and accountability goals and objectives and the assessment instruments that would be used to achieve them and measure progress. The plan could include specific programs for in-service training, public information, and for assuring that tests are selected, used, and interpreted appropriately.

Finally, we do not believe that the federal government should get into the state and local testing business. We commend the efforts to develop national education goals, and stand ready to develop and administer the assessment instruments required to measure progress toward attainment of such goals.

I would be remiss if I did not point out that while publishers are trying to respond to the need to develop challenging and innovative tests (parallel efforts are being undertaken by publishers of textbooks and other instructional materials), federal tax policy is frustrating its achievement.

The Department of the Treasury is insisting that publishers of tests and instructional materials capitalize research and development and other pre-publication costs, a position that falls with special weight on preparation of new tests and instructional materials, with their high development costs, high risks, and long lead times. This approach is shortsighted as a matter of educational policy because it discourages the development of the innovative quality tests and textbooks our schools need. It is also discriminatory and unjustified tax policy because it requires capitalization of product development and research costs that, for any other industry, could be deducted in the year incurred. We have requested the tax-writing committees (and the Administration) to provide appropriate relief, but the outcome remains very uncertain. This Committee's assistance in assuring that tax policy does not frustrate education policy would be most welcome.

Thank you for your attention. I would be pleased to respond to any questions the Committee may have.

1779h

EDUCATIONAL MEASUREMENT
ISSUES AND PRACTICE
Volume 7, Number 2
Summer 1988

Riverside Comments on the Friends for Education Report

Edward C. Drahozal
Riverside Publishing Company
and
David A. Frisbie
The University of Iowa

The authors, both affiliated with Riverside Publishing Company, discuss the factors they think explain the Lake Wobegon phenomenon and call for more appropriate use of normative comparisons and more complete reporting of test results.

The final report of Friends For Education (FFE), "Nationally Normed Elementary Achievement Testing in America's Public Schools," has appeared in several preliminary forms, and portions of its contents have been discussed in several national newspaper reports. The following comments are based solely on the content of the final report, and, though we take exception to some of the data analysis methods employed, some of the interpretations offered, and some of the conclusions reached, these remarks focus primarily on questions raised by the report. These are issues that complicate the use of aggregated test scores—with regard to both their formation and their interpretation—to synthesize state versus national comparisons in the face of different test batteries, different standardization samples and methods, different years of norming, and different score scale units for reporting.

The study of state and district performance reported by FFE appears to have been conducted as carefully as possible under the circumstances. The issues that the report raises—accuracy and comparability of norms, currency of norms, selectivity of pupils tested and reported in reports to the public, and temptation to teach specific content when educators are under accountability pressure—are not new. What comes as a genuine, unexpected, disappointing shock is the apparent universal appeal of the

---

Edward C. Drahozal is Senior Program Administrator at The Riverside Publishing Company, 8420 Bryn Mawr Avenue, Chicago IL 60631. He specializes in practical aspects of interpretation, use, and reporting of elementary and secondary tests, primarily standardized tests.

David A. Frisbie is Associate Professor and Assistant Director of the Iowa Basic Skills Testing Program, University of Iowa, 311 Lindquist Center, Iowa City, IA 52242. He specializes in achievement testing.

simplistic objective of being above the national average and the extent to which schools are successful in managing somehow to appear above the national average when faced with pressures and even ultimatums from politicians, press, the courts, and even watchdog groups.

Despite the shortcomings that can be cited regarding the nature of the data that FFE analyzed and reported, there is ample evidence to warrant close examination of the group's fundamental question: Why are so many pupils or schools or states appearing to perform above the national average? The question seems as simple and straightforward as the one posed several years ago: Why are test scores declining from year to year? We believe that the question raised by FFE rivals the score-decline question in significance and, as was true of the score-decline inquiry, this search for resolution is likely to yield multiple, concomitant explanations. There is no single best answer. A closer examination of the issues by FFE, the publishers, and the state and district test coordinators might enhance our ability to use test data to further our primary goal—to improve the quality of instruction provided in our schools. With this purpose in mind, the remainder of this paper is devoted to identifying what we believe are the most crucial issues and to presenting a scheme that we would use to compare the performance of state or district groups with national pupil or school norm groups.

Some Major Issues

Accuracy of Norms

National norms for standardized achievement tests are based on a sample of pupils and schools (attendance centers) obtained through a complex, multistage sampling scheme. Each publisher strives to ensure that the national population of pupils and schools is properly represented in its norms sample. For example, in the standardization of the Iowa Tests of Basic Skills (ITBS) in 1984–85, districts were chosen on the basis of geographic region, enrollment size, and socioeconomic characteristics of the communities served. The standardization is a joint responsibility of the authors, publishers, and school personnel. Rigid conditions for district participation included the provision for sampling attendance centers of the district by the publisher rather than by the school administration. An adequate sampling plan is necessary but not sufficient to guarantee adequate norms. Only if the plan is realized, only if the sample obtained reflects the sample desired, will the norms represent national pupil or school achievement accurately.

To the extent that any publisher's norms misrepresent the national distribution of pupil and school achievement, comparisons with either of these norm groups will distort the estimated achievement level of the group in question. An underrepresentation of high-achieving schools or high-achieving pupils will cause the national norms to be "softer" than they ought to be. That is, an average-achieving pupil will appear to be above average when referenced to a group whose average is below their theoretical or "true" average.

The sampling plan, nature of the obtained sample, and weighting schemes used in the standardization of each achievement battery in question should be examined to determine the representativeness of the published norms. This should be done separately for pupil and school norms.

Recency of Norms

It is a well-documented fact that achievement in grades 3–8 has been rising steadily since the late 1970s. Though the year-to-year differences might be regarded as minor (.3 of a grade-equivalent month, on the average), the cumulative effect over 10 years is significant (approximately 3 months, on the average). Obviously, those who compare the 1987 performance of their pupils with that of other pupils who were tested in 1978 (national standardization) will be using "softer" norms and will have more pupils appearing to be above the national average than really are.

We have published information on changes in student performance for the past 30 years. Data for 1955 to 1984 are summarized on pages 148–153 of the new ITBS *Manual for School Administrators* (Hieronymus & Hoover, 1986). Differences in performance vary by test, grade, and score level. The 1977–85 composite score differences are eight to nine percentile ranks (PRs) at the median in *most* grades, but differences in language exceed 10 PRs in several grades at several score levels.

In periods of fluctuating achievement levels, the recency of the norms is a critical issue. When achievement levels are relatively stable over time, as they have tended to be at the grade K–2 levels, "old" norms do not interfere with score interpretations, assuming that we have curriculum stability as well.

Nature of Tested Population

If we have good reason to believe that pupils in a given state should have scores, on the average, below the national average, we must be certain to define the population for which we expect the prediction to hold. There are several related issues regarding this point with respect to the FFE data. If State X reports a mean normal curve equivalent (NCE) for 45,000 fourth graders, we should ask these questions: How many fourth graders were tested but not included in the computation of the mean, and what is the nature of the scores of those who were excluded from reporting? How many fourth graders are there in State X who were not tested and, consequently, who were not included in the reported scores? And what are the achievement levels like for these students who were not tested?

Based on the Department of Education's Center for Education Statistics fall 1985 enrollments projected to 1986, the percentage of students for whom scores are reported in the FFE report varies from a low of about 85% to more than 96% of total grade enrollments for most states for which full-grade testing was reportedly done. (For one state with public school enrollments of about 48,000 students per grade, averages and PRs are reported for approximately 37,500 students, which is about 80% of the total enrollment.) The discrepancy between the reported state scores and the expectations in the FFE report may be in part due to such differences between tested and total enrolled populations of students and specifically to the nature of the portion of the student population not tested.

Adequacy of Expectations

Educators have developed some expectations about how pupils and groups of pupils might perform on achievement tests based on their study of the relationship of school achievement to other social, political, and economic variables. This is why we use such variables as enrollment size and socioeconomic status for stratified sampling in standardizations. FFE has used some of these relationships in attempting to develop expectations for state-level and school-district-level performance. Per-capita income, graduation rate, and college entrance score averages are among the "standard barometers of excellence" employed by FFE. Though we do not deny the value of these indicators as part of the prediction equation, we realize that it is not possible to predict achievement in this way with high accuracy. For example, the achievement test performance of Iowa pupils is among the very highest in the nation, yet these facts about educational conditions in Iowa seem inconsistent with that high level: Iowa ranks 27th among states in per-pupil expenditure, 39th in average teacher salary, and 44th in spending increase from FY86 to FY87.

In view of the less-than-perfect relationships between achievement and these other variables, the precision of whatever expectations about achievement we may formulate should be tempered. That is, what we are able to say with reasonable assurance about how many pupils or schools should score above a specific point (the mean in the national norms distribution) is not very useful. Consequently, we might instead settle for statements like these for State X: "About 40% of the fifth graders tested should score between the 25th and the 75th percentiles on national pupil norms," or "About 49–55% of the third graders tested should score above the national pupil median (50th percentile)." Of course, the ability to make such statements depends on a far greater understanding of the statistical relationship between those variables than most states probably have been able to determine.

Teaching the Test

Pupils and their teachers who participate in the standardization of an achievement battery have not had an opportunity to see or study the specific test questions used. Thus, having no practice on the specific test questions is one of the stringent preconditions of the standardization process. Subsequently, when these norms are used to interpret the scores of pupils who have been drilled with the exact test questions, the result is an overrepresentation of the amount of knowledge and skill possessed by such pupils. Likewise, when the scope of the curriculum is narrowed to encompass primarily the objectives measured by the exact test questions, the relative standings of the pupils who experienced the restrictive program of study will be overestimated.

No publisher condones this use of tests, and few teachers probably follow such abominable practices. Those who do are nearly always motivated by significant negative consequences associated with scores that might turn out to be below expectation (not always synonymous with national average). Unfortunately, for some educators, job retention and salary increases are tied directly to the test scores of their pupils. The authors of the ITBS have always decried the use of achievement scores for such purposes and instead have campaigned for the use of these scores to improve instruction directly.

If certain tests are to be used strictly for accountability purposes, their security must be ensured so that the scores that result will be valid for that purpose. The dollars required to assure states and districts that the test forms they will use are secure would be far greater than the value of the information derived from using the secure forms. Those dollars would likely have greater and more visible impacts on learning if devoted to direct instruction instead.

Score Analysis and Interpretation

With which norm group, pupils or schools (attendance centers), should averages from State X be compared to interpret the scores of pupils from that state? With which norm group, pupils or schools, should averages from District A be compared? There are only two choices, pupils and schools, because no publisher provides norms for school districts or for states. This is a fundamental issue currently facing the Council of Chief State School Officers as they contemplate options for providing for state-by-state achievement comparisons in the future. The choice to be made is not a matter of personal preference but a matter of the logical correspondence between the units to be compared. That is, averages of school buildings should not be referenced to a distribution of individual pupil scores, district averages should not be referenced to the distributions of either school building averages or pupil scores, and state averages should not be referenced to any of these three distributions. In view of the differences between these separate distributions, it is most logical to reference a score or average score to its own kind. When the most logical referencing is not possible, appropriate caution should be exercised.

The national pupil norm group includes pupils whose scores on a test are as high as perfect (PR = 99) to those whose scores are as low as zero or chance average (PR = 1). No school (building or attendance center) is likely to have an average score that is perfect or zero. In fact, on the ITBS and any other test with recent school norms it is reasonable to expect that no school will have a raw or scale score average higher than PR 88 or lower than PR 12 compared to the pupil distribution. Because many school districts are single-grade-within-single-building entities, the distribution of school district averages probably would encompass the same range as the distribution of school building averages. The school district distribution, however, is likely to be markedly more leptokurtic and less variable than the school average distribution. In terms of the pupil distribution, the distribution of district averages might range, *effectively*, between PR 75 and PR 25. Finally, *most* of the state averages on a test for a given grade might well have actual bounds that corre-

TABLE 1

Percentages of State X Pupils Performing Within Selected National Pupil Percentile Intervals

National percentile rank	National percentage	Grade K	1	2	3	4	5	6	7	8	Average K-8
90-99	10	20	21	24	20	21	20	20	21	19	20.7
75-89	15	22	24	23	24	23	25	24	24	23	23.5
50-74	25	26	29	25	27	28	28	29	28	28	27.4
25-49	25	19	18	18	17	16	17	17	17	19	17.6
10-24	15	8	6	7	7	8	6	7	6	7	6.9
1-9	10	5	2	3	4	3	3	3	4	4	3.4
Percentage above national median		68	74	72	71	72	73	73	73	70	72
Percentage below national median		32	26	28	29	28	27	27	27	30	28

spond to PR 60 and PR 40 on the pupil distribution.

Because norms for district averages or for state averages are not available, districts and states often use the pupil and school norms that do exist. When a district average is referred to the pupil norms, it should be thought of as the score obtained by the average pupil in the district. We might find, for example, that the average pupil in District A scored higher than 63% of pupils nationally. Using the same rationale and the estimate given above, the average pupil in most states is not likely to exceed PR 60 or fall below PR 40. The value of such information is highly questionable.

A matter related to this general issue of analysis concerns the methods of computational precision used to aggregate and convert scores. As an example of the problem, a grade 4 school average GE composite score of 42.0 (obtained in the fall) on the ITBS has a PR of 46, and a score of 43.0 has a PR of 53. By interpolation and rounding, an average GE of 42.5 corresponds to a PR of 49.5 or 50. If GEs are rounded before converting to PRs, a 42.5 could be treated as a PR of 46 or 53, depending on the rounding convention adopted. Of course, this illustration underplays the magnitude of the distortion that could result with distributions of either school district or state averages.

Publisher-User Responsibilities

Though it is in the best interest of both publishers and test users to have tests and scores used properly, neither can ensure that the other will do its part willingly and unselfishly. Publishers must be counted on to standardize and analyze results in professionally acceptable manners. They must guard against potential misuse by informing educators of the intended uses of the tests they publish and warn against the possible misuses that might be anticipated. Publishers must do their utmost to provide test materials only to those who are at least minimally qualified to handle the tests and scores in a professional way. State directors, superintendents, teachers, school boards, and the public, generally, do not have the resources to monitor the effectiveness of publishers in attending to these obligations.

Publishers, on the other hand, cannot monitor the use of their instruments effectively to curtail misapplication, misuse, or misinterpretation. Often after the fact, a publisher can recognize inappropriate use—whether intentional or unintentional—and attempt to persuade the user to modify a proposal or report. Some school districts perform extensive audits to ensure that all students who were to be tested in each attendance center were actually tested. Some districts also audit results *and* retest suspect groups. But for the most part, publishers are not aware of and have no control over school districts' test administration conditions, the students included in summary data reported to the public, or methods used to synthesize data to make test results more palatable for less sophisticated consumers.

Most test authors and publishers go well out of their way to comply with the standards for educational and psychological tests adopted by the profession. Test score users—teachers, administrators, legislators, and other public groups—tend to know far less than they should about the nature of tests or the principles with which test makers intend for scores to be used. We should not denounce a test because a state committee uses the wrong norms or incorrect statistical analysis procedures in reporting. Likewise, we should not blame users for results based on shoddy standardization procedures or on inadequate or deceptive descriptions of such procedures.

Finally, publishers are obligated to clients to maintain the confidentiality of test data. It has been and should continue to be each client's decision to release test data and to determine the nature of any data to be released. Reporters, citizens, citizens' groups, and others who wish to obtain test data should respect this publisher-client relationship and seek release from the school district or state, depending on their level of interest and the dictates of state law.

A Sample Reporting Method

We recommend an approach like the one described below for states that wish to describe the achievement levels of their pupils in relation to pupils in a nationally representative norm group. Exactly the same procedures could be used with school building (attendance center) data. Table 1 shows national PR ranges in the first column and the corresponding percentages in the second column. The body of the table shows, separately for each grade, the percentage of pupils in State X that obtained national PRs in each range. The last column shows the row averages of the percentage values. (Note that these are percentages and not percentile ranks and, consequently, it is acceptable to average them.) The bottom two rows indicate, again by grade, the percentage of pupils above and below the national median. A histogram with one distribution superimposed on the other or a simple bar graph would provide a helpful visual display of the same information.

The main advantage of this method of reporting compared with reporting simply the percentage scoring above the national median is obvious. Between-grade differences and similarities can be examined, but most important, discrepancies from the national distribution can be accounted for in each of several segments of the distribution. If all we know is that 72% are above the national median, we do not know if the "extra" 22% are mostly located very near the median, mostly spread through the upper half, or mostly concentrated in the tail. Also, we do not know if the extra 22% are shifted from the lower tail, from throughout the lower half, or from just below the median.

Many districts use a reporting procedure similar to the reporting scheme described above. We recommend that such tabular data be supplemented with at least the following sorts of information: testing date, test form and level(s) used, type and date of the norm used, and percentage of eligible students tested.

Riverside Publishing Company and its representatives do not believe that the average pupil in every state has scores above the national median on the ITBS. We are confident in our standardization procedures and have subjected those procedures to public scrutiny in detail in the *Manual for School Administrators*. We have updated our norms at least every 7 years and, when achievement showed a pattern of increase in the early 1980s, new norms were obtained even though new test forms were not introduced. We are making plans to provide annual national norms updates for next year. Our manuals caution users about appropriate use of norm groups for varying purposes. Our hope is that the issues raised above will cause FFE and state and district test coordinators to reassess their analysis and reporting procedures to ensure that conclusions reached are based on a valid foundation rather than data of questionable origin and manipulation.

References

Hieronymus, A. N., & Hoover, H. D. (1986). *Manual for school administrators. Iowa Tests of Basic Skills*. Chicago: Riverside.

A Response to John J. Cannell

Joanne M. Lenke and John M. Keene
The Psychological Corporation

Two representatives of The Psychological Corporation present their reactions to the Cannell report and call for better explanations for the public of the meaning and limits of norm-referenced scores.

In recent years, public attention has focused on standardized achievement test results. These results, which are intended to describe the performance of individuals in relation to one another, are now often used to describe the performance of groups of students. In a report entitled "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All 50 States Are Above the National Average," John Jacob Cannell attempts to cast doubt on the validity of the information being reported to describe the achievement of students as aggregated at the state and/or district level. The report states, "These tests allow all the states to claim to be above the national average! The tests . . . allow 90% of the school districts in the United States to claim to be above average. More than 70% of the students tested nationwide are told they are performing above the national average."

In response to Cannell, it is fair to say that many states and school districts report above-average performance in reading, mathematics, and/or language in the elementary grades. We do not believe that this is an attempt to misrepresent students' achievement in the nation's schools. Let us examine three very

---

Joanne M. Lenke is Vice President, Measurement, at The Psychological Corporation, 555 Academic Court, San Antonio, TX 78204-2498. She specializes in test development and norming, scaling, and equating tests.

John M. Keene is Director, Admissions/Credentialing, Customized Measurement, and Research, at The Psychological Corporation, 555 Academic Court, San Antonio, TX 78204-2498. He specializes in educational measurement.

important issues related to the interpretation of this information: (a) group performance relative to a national norm, (b) local performance relative to national performance, and (c) the stability of achievement test norms over time.

Interpreting Group Performance Relative to a National Norm

When a test is standardized, or normed, the test is typically administered to hundreds of thousands of students nationwide. This norming sample is drawn to reflect specified demographic characteristics of children attending school in the United States. Such demographic characteristics include socioeconomic status, ethnicity, region of the country, and size of school district. Percentile ranks are then derived from frequency distributions of individual students' scores at each grade. Norms provide a mechanism for describing a student's performance relative to that of other students in the same grade from across the country at a particular point in time.

The use of these norms to describe group performance must be interpreted carefully. For example, if a state's average score in reading is at the 54th percentile, the proper interpretation of this score is that the average, or typical, *student* in the state performed better than 54% of the norming sample. It is not appropriate to conclude that all students in the state are above average in reading, that the state as a whole is above average in reading relative to other states, or that the state as a whole is above average in reading relative to the national norm.

The approach used by some states and school districts in the reporting of group performance is to report the percentages of students scoring, say, "at or above the 50th percentile," or "in the average and above-average range." Although this method of reporting is appropriate because it maintains the relationship between individual performance and the national norms, the reported percentages should be accompanied by corresponding percentages for the national norming sample. Although it is obviously the case that 50% of the national sample of students scored at or above the national median at the time the test was standardized, it may not be the case that 50% of the national sample scored at or above the national mean raw score or national mean scaled score. If the reporting metric is something other than the percentage of students scoring at or above the national median, the appropriate national comparison should be provided so that proper inferences about the data can be made.

Interpreting Local Performance Relative to National Performance

It is unlikely that the demographic characteristics of the students in any state or school district mirror those of the nation as a whole; it is equally unlikely that the curriculum of any state or local district is as diverse as that of the nation as a whole. Furthermore, it is not necessarily the case that the guidelines set forth by the test publisher with regard to the testing of handicapped or limited-English-proficient students in a norming program are the same as those used in actual practice. If there were a state or district whose demographic characteristics matched those of the nation, whose curriculum was as diverse as that of the nation as a whole, and whose administration guidelines and procedures were consistent with those used by the publisher for the norming sample, one would expect the average student in the group to score at about the 50th percentile. To the extent that differences exist, we must remind ourselves that when local group summary scores are interpreted in reference to a national norm, the interpretation has to be placed in the proper context, simply that of the group's average student relative to the national norm. Because it is unlikely that the students tested in any given state or district are typical of the nation in all respects, it would be unreasonable to expect *any group* to be *at* the national average.

Test purchasers, districts as well as state agencies, often select tests through a process that examines the match between the test content and the local curriculum. In many cases, the selected test is the one that best reflects the local curriculum. Test users selecting tests on this basis may have an advantage over the norm group because the test is likely to be more valid for assessing performance in the local curriculum than it is for assessing the performance of a national sample of students being exposed to different curriculums, presumably having somewhat different emphases.

The Stability of Achievement Test Norms Over Time

Cannell's report suggests that the use of "old" norms is partially responsible for high achievement test scores. Presently, test publishers produce new editions of their tests on a 7-to-9-year cycle, and current norms are provided with each new edition. Because test adoption cycles do not necessarily coincide with test revision cycles, it is conceivable that the norms for a newly adopted test may be 2 or more years old. Therefore, it is critically important that empirical norming dates accompany the reporting of achievement test results.

It is very encouraging to note that today's students are performing better than their counterparts did in the late 1970s and early 1980s. Evidence of this improvement in performance can be found not only from research that test publishers have conducted in equating newly published tests to previous editions, but also from a recent research study conducted by The Psychological Corporation with the current edition of the *Stanford Achievement Test* Series. First standardized in the 1981–82 school year, the Stanford Series was administered to a nationally representative sample of 350,000 students in spring and fall 1986. The sample was further stratified according to "user" and "nonuser" groups, where "users" were defined as school districts that had been using the Stanford in one or more grades for at least one year in their districtwide or statewide assessments. The results of this study revealed that "users" outperformed "nonusers," and, more importantly, that "nonusers" performed better than the original norming sample in mathematics, reading, and the language arts in the elementary grades. Two important generalizations can be made from this research. First, test scores do tend to increase when the same test series is used year after year. However, this should not necessarily be attributed to "teaching to the test"; rather, the test results provide a needed focus on areas in need of improvement. Second, educational achievement did improve from 1982 to 1986 in some subject areas in the elementary school grades. Therefore, more current norms for the Stanford Series have been developed and are available to users of the battery.

During this time of educational improvement, it is important not to lose sight of the fact that use of the same norms over a period of years enables the test user to demonstrate improvement relative to a constant reference group. Even if it were economically feasible for test publishers to produce representative national norms more often, frequently updated norms represent a "moving target," where educational gains (or losses) would be masked by the relative nature of the information. The level of achievement of students in the United States has increased in recent years, and educators must have the opportunity to demonstrate these gains in order to ensure the necessary support of the local community in improving the quality of education. The education of young people must continue to improve, and norm-referenced achievement tests are useful tools in this endeavor.

Conclusion

Because the public is expecting norm-referenced scores to represent standards of performance, we, as educators, must assist the public in becoming better informed about the interpretation of test results. National normative data provide extremely important information for making sound educational decisions. The degree to which these decisions are defensible depends on a clear understanding of the strengths and limitations of the data.

The Time-Bound Nature of Norms: Understandings and Misunderstandings

Paul L. Williams
CTB/McGraw-Hill

Presenting a view from CTB/McGraw-Hill, the author discusses the time-bound nature of test norms and argues that the phenomenon of most elementary students' scoring above averages from previous years' norms is a result of generally increasing levels of achievement.

Recent interest in the topic of the time-bound nature of normed scores has resulted, in part, from allegations made in a report issued by the Friends for Education. The key element of the argument put forth in the Friends for Education report is that too many students appear to exceed the national average. Data have been presented in the report which are said to show that more states and school districts are scoring above average than one might initially expect.

It is an interesting phenomenon that it is through the vehicle of the Friends for Education report that the time-bound nature of norms has received some measure of public attention. The fact that norms have always been referenced to the year of test standardization is something that has been so universally known and understood by testing professionals that it has not had a large measure of attention focused on it. Perhaps that will prove to be an important singular contribution of this issue of *Educational Measurement: Issues and Practice*.

The Cyclical Nature of Test and Norms Development

The evolution of norm-referenced tests (NRTs) as valuable assessment instruments has been characterized by the expansion of the purposes for testing. In the earlier versions of NRTs (in the mid-1960s to the mid-1970s), the primary purpose was to provide accurate normative scores so that group and individual comparisons could be made to a national profile of achievement. Using this information, school adminis-

---

Paul L. Williams is Director of Research and Measurement at CTB/McGraw-Hill, 2500 Garden Road, Monterey, CA 93940. He specializes in educational testing and measurement.

trators could evaluate programmatic and individual strengths and weaknesses so that appropriate instructional intervention and resource allocation could be applied. Additionally, using multiple-year testing, longitudinal trends in achievement could be monitored.

An expansion of these purposes took place with the publication of the *California Achievement Test* (CAT), *Forms C and D* (CTB/McGraw Hill, 1977). This test battery, for the first time, allowed scores for instructional objectives to be reported from an NRT for individual examinees. Although earlier NRT test versions did allow test administrators to use item analyses for minimal diagnostic purposes, CAT C and D provided specific instructional objective scores for the purpose of more individualized instructional planning.

The schedule for the publication of norm-referenced tests has followed a basic, industrywide cycle of between 5 and 8 years for the same test series. In the instance where a test company has more than one NRT series, such as CAT and the *Comprehensive Test of Basic Skills* (CTBS), publication is staggered so that one test of the series is published about every 3 or 4 years.

This cycle has been dictated by several factors. The first factor has been the speed with which curricular changes take place in the nation's schools. NRTs are designed to reflect the predominant achievement outcomes and curricular trends in the nation's schools. When a new form of an NRT is developed, content considerations are of paramount importance. Although curricular trends have a major impact on the content of NRTs, these trends do not change so fast in the schools that more frequent revisions of a test series would be justified based solely on them.

At the time an NRT is revised, the collection of data for the generation of new national norms takes place. Using a national probability sample, data are collected for several hundred carefully selected school districts and hundreds of thousands of students. Based on this carefully selected stratified sample, normative scores are developed.

Each of the derived scores that emerge from the standardization process, including percentile ranks, grade equivalents, and normal curve equivalents (NCEs), has a predefined relationship to the characteristics of the norm group. Thus, at the time the test is normed, 50% of the examinees will exceed the 50th percentile and the same percentage will fall below the 50th percentile. Derived score tables for the test battery are produced, and all scoring of student tests is referenced to these tables until the battery is either revised or, in rare instances, when it is renormed with no change in the content of the test.

Data from national probability samples are not usually collected for a test more often than every 5 to 8 years because it is impractical and economically infeasible to do. It would not be reasonable to ask or expect schools to administer tests to large numbers of students every school year in order to develop yearly norms based on a national probability sample. The cost of such testing would have to be passed on by the publisher to the schools and would add substantially to the cost of school testing programs.

In summary, most large test publishers follow the common and decades-old industry practice of revising and standardizing their achievement tests about every 8 years. The content is updated to reflect current curricula and instructional practices, and new norms are developed so that the test reflects levels of achievement that prevail during the school year in which the test is standardized. The dates of standardization are given wide publicity, and all purchasers of the test are aware of these dates.

Proper Interpretations of National Norms

Because norm-referenced tests are not normed yearly on a national probability sample, changes in national achievement between the norming years will be reflected in the norm scores for groups of students. For example, if national achievement levels decrease between normings, as they did from the late 1960s to the mid-1970s, students' norm-referenced scores will decrease, and more students will fall below the median (50th percentile) score established when the test was normed. On the other hand, when national achievement levels increase between normings, more students will exceed the median established when the test was originally normed. Regardless of the direction of national achievement trends, when a test is renormed, exactly half of the students will fall above and half will fall below the newly established median.

At this time, national achievement indicators all point to the fact that student achievement is generally on the increase. This increase is documented by the National Assessment of Educational Progress (NAEP); the Scholastic Aptitude Test (SAT) results; two Congressional Budget Office reports (1986, 1987); and data collected during recent test normings by CTB/McGraw-Hill (1985, 1987, 1988).

Thus, during a time of increasing national achievement, the students' normed test scores will rise between norming periods. More students will score above the median score established during norming than will fall below it. This confirms the sensitivity of the test norms to changes in achievement, one of the tests' primary functions. These normed test scores are valid measures of student growth. Although the reference year for the scores will be prior to the year in which the test scores are reported, the test scores provide accurate program and student information. The fact that the norm scores themselves refer to norming that took place during an earlier year in no way compromises the major purposes for administering an NRT or the usefulness of the scores for program evaluation, student instructional planning, or the monitoring of longitudinal trends. When interpreting the scores, the test user must simply be aware of the year that the tests were normed and the general direction of national achievement trends. Interpretive guidelines are found in relevant test-related materials produced by most publishers.

The Friends for Education report has received attention primarily as a result of its improper interpretations of score distributions for always provided.

The second example relates to the Annual National Normative Trend Data (NTD) published by CTB/McGraw-Hill. Research on this project began in 1984, when an emerging customer need was identified by the company. Customer comments about the desirability of obtaining more recent normative data were noted in market research efforts. Such data could be used to amplify the standardization norms and provide a more complete picture on the progress local school districts were making in their instructional efforts. After 3 years of research, the NTD service was offered to CTB customers. Score reports have been made available on an annual basis, for the standardization year as well as for the most recent norming. This service is a response to those educators who have been concerned about the time-bound nature of norm-referenced scores.

The test companies do their best, through many vehicles, to assist the test consumer in being a responsible user of test results. Indeed, reasonable testing programs, effectively implemented, are one of the reasons that achievement is increasing and that we are not currently in the decline phase that manifested itself in the late 1960s to the mid-1970s.

The assertion that scores are on the increase does have merit. Perhaps the positive side of this phenomenon should be stressed more. States and local school districts have committed considerable resources to improving the achievement levels of their students. All indicators of student achievement appear to converge on this fact, particularly for the elementary grades. The American public should be gratified that achievement is increasing.

Cannell (1987) charges that "inaccurate initial norms and teaching the test," rather than improved achievement, are reasons for improving scores on nationally normed tests. The problem with these allegations is that there is little, if any, evidence to support them. To the contrary, the body of independent evidence suggests that test norms provide a valid and useful reference in both the norming year and in subsequent years and that achievement at the elementary level has been increasing. If indeed there exist instances of abuse of test norms and of misunderstanding of their meaning by educators or the public in general, then the proper remedy should be to correct those instances rather than to make rash allegations about the adequacy of test norms or questionable teaching by educators.

References

Cannell, J. J. (1987). *Nationally normed elementary achievement testing in America's public schools: How all fifty states are above the national average.* Daniels, WV: Friends for Education.

Congressional Budget Office. (1986). *Trends in educational achievement.* Washington, DC: Author.

Congressional Budget Office. (1987). *Educational Achievement: Explanations and implications of recent trends.* Washington, DC: Author.

CTB/McGraw-Hill. (1977). *California Achievement Test, Forms C and D.* Monterey, CA: Author.

CTB/McGraw-Hill. (1981). *Comprehensive Test of Basic Skills, Forms U and V.* Monterey, CA: Author.

CTB/McGraw-Hill. (1985). *California Achievement Test, Forms E and F.* Monterey, CA: Author.

CTB/McGraw-Hill. (1987). *Annual national normative trend data.* Monterey, CA: Author.

CTB/McGraw-Hill. (1988). *Annual national normative trend data.* Monterey, CA: Author.

SRA Response to Cannell's Article

Audrey L. Qualls-Payne
Science Research Associates

The author defends SRA's norms, discusses some of the difficulties in pursuing Dr. Cannell's proposals, and points out that we need to monitor not just student achievement levels but also trends in curriculum.

Science Research Associates (SRA) recognizes the concerns expressed in John Cannell's article, "Nationally Normed Achievement Testing in America's Public Schools: How All 50 States Are Above the National Average." We differ, however, in our assessment of the situation and the proposed alternatives. According to the article, most schools in the nation perform at or above average on commercially available tests. This finding, as noted by Dr. Cannell, is not consistent with statistical theory, which says that half the students should be above and half below. Dr. Cannell expresses the opinion that this inconsistent statistical phenomenon results from using older tests, older norms, teaching to the test, statistical manipulation of the data by publishers, excluding special

---

Audrey L. Qualls-Payne, Psychometrician, Science Research Associates, Inc., 155 N. Wacker Dr., Chicago, IL 60606-1780. She specializes in achievement testing.

education students from the calculation of group averages, and inaccurate norms. He goes on to suggest that these problems can be eliminated by the use of one achievement test in all schools across the country with the concurrent development of annual norms. Our purpose is to examine Dr. Cannell's conclusions and offer alternatives to some of the issues raised in his report.

SRA's national norms are reliable and accurate indicators of national student performance at the time of standardization. The charge of statistical manipulation of data appears to result from Dr. Cannell's apparent misunderstanding of the purpose of the various types of test scores and subgroup norms. Schools may wish to compare their students' performance with, in addition to that of the national group, that of groups more similar in structure and student composition. For example, a nonpublic school may want to compare their students' performance with that of students from other nonpublic schools. The various test scores, in addition to status scores (i.e., percentiles and stanines), are offered to meet the many needs of our customers. Normal curve equivalents (NCEs) are required for Chapter 1 program evaluation. To assess longitudinal growth and determine functional levels, developmental scores, for example, standard scores and grade equivalents, are needed.

Dr. Cannell's alternative to the various standardized achievement tests is a national achievement test, which would require at least two major actions. First, this national achievement test would have to be normed annually with a representative group of students to have yearly norms. Second, new test forms would be needed for each administration to eliminate possible problems of teaching to the test and test security.

A project of this magnitude and complexity would be very difficult logistically and very costly. Two major logistic problems would be (a) obtaining curricular consensus on the test content and (b) obtaining or mandating national participation.

If yearly new forms are not an option but annual norming is, and if there truly is a substantial amount of teaching to the test, the problems noted in Dr. Cannell's analysis may not go away. If new forms of achievement tests are developed each year, thereby increasing test security, the need for annual norms diminishes significantly. Based on Dr. Cannell's analysis from schools with tight test security and literature on student growth, drastic shifts in student performance from one year to the next are rare. From a psychometric point of view, new norms are needed only when there is a *significant* shift in school curriculum and/or student performance.

As opposed to developing and standardizing new forms each year, a mechanism is needed to monitor changes in school curriculum and student performance. Whenever there is a change in either curriculum emphasis or achievement levels, new test forms should be developed and standardized. If the change is strictly a shift in student achievement, renorming is required. As a publisher, we must base our decision on when to issue new forms/new norms on a systematic monitoring system.

There are several ways to monitor student progress. One way to accurately spot when significant changes are taking place is to track student achievement on a regular basis (i.e., annually). The entire user group could be used for this purpose. The monitoring process should be capable of producing user-based norms, which can then be made available to all customers as an optional service in addition to the national norms.

There is at least one major problem with the user-based monitoring system. If the user sample is biased and unrepresentative of the national student population, significant changes noted in the user sample may not truly reflect changes at the national level. One way to resolve this problem would be to select a subset of schools from the user group and use it to monitor changes in curriculum and student achievement annually. The selected schools should be representative of the national population of schools with respect to geographic region and racial/ethnic and socioeconomic status. Once a set of schools is selected for this purpose, students in these schools can be tested on an annual basis and norms can be developed. As in the previous method, annual norms will be made available to customers as an optional service. Because of the representativeness of the schools selected for monitoring, one can, with a high degree of confidence, generalize results from this set of schools to the U.S. population of schools.

Because SRA recognized the value of a monitoring system, we are already in the developmental stages of implementing such a program.

MIDDLE CITIES EDUCATION ASSOCIATION
517 Erickson Hall • Michigan State University • East Lansing, MI 48824-1034
(517) 355-1720 • FAX (517) 353-6393

OFFICERS
President
James Agee, Muskegon

Directors
Larry Mueller, Port Huron
(President Elect)
Ira Rutherford, Beecher
Richard Benjamin, Ann Arbor
Richard Halik, Lansing

Executive Director
Michael A. Boulus

MEMBER DISTRICTS
Adrian Public Schools
Ann Arbor Public Schools
Battle Creek Public Schools
Bay City Public Schools
Beecher Comm. Schools
Benton Harbor Schools
Buena Vista School District
Flint Comm. Schools
Grand Rapids Public Schools
Highland Park Public Schools
Jackson Public Schools
Kalamazoo Public Schools
Lansing School District
Midland Public Schools
Monroe Public Schools
Mt. Clemens Comm. Schools
Muskegon Public Schools
Muskegon Heights Public Schools
Niles Comm. Schools
Plymouth Canton Comm. Schools
Pontiac School District
Port Huron School District
Saginaw Public Schools
Traverse City Public Schools
Waterford School District
Willow Run Comm. Schools
Ypsilanti School District

March 19, 1991

The Honorable Dale Kildee, Chairman
Subcommittee on Elementary, Secondary
and Vocational Education
320 Cannon Building
Washington, D.C. 20515

Dear Congressman Kildee:

Please accept this letter regarding the development of a national assessment system as input from the Middle Cities Education Association for the record of the March 14, 1991, Oversight Hearing.

The Middle Cities Education Association (MCEA) is a consortium of 28 central city school districts in Michigan which serve 300,000 (20%) of the students in the State. The organization works to create a better understanding and support for the needs of urban school districts and the many economically and educationally disadvantaged children we serve.

While it is clear that there is renewed interest in and movement toward a national testing system, we believe there are a number of issues which need to be addressed. The following issues were discussed recently by the Directors of Testing and Evaluation in the MCEA school districts:

- Any effort toward a national assessment should have a clearly defined purpose. We are uncertain whether a national test is being proposed in order to provide our nation's students with essential knowledge and skills, or whether the movement is more political in nature. We would hope that the need to increase student performance by improving curriculum and instruction is the driving force behind national testing.
- Prior to adoption of a national assessment, a well developed national curriculum framework needs to be developed so that teaching and learning for all children is based on a common

set of clearly stated goals, outcomes and standards. Within a broad curriculum framework which describes student outcomes to be achieved, states, school districts, schools and teachers should be free to creatively employ a variety of instructional strategies and materials to ensure students achieve those outcomes. Development and use of an assessment without first having this curriculum framework would be measuring apples and oranges. An every-pupil, nationwide test is meaningless if what is being taught differs widely. In other words, a test which measures progress toward previously specified common goals and student outcomes would be meaningful and would not necessarily stifle local creativity in development of the means to reach those outcomes.

- Any national test should be based on a rich, challenging curriculum that goes beyond basic skills and rote learning. We are concerned that a national test/curriculum may stall a needed shift toward process skills. Today's students need to demonstrate the ability to think critically, solve problems, apply knowledge, utilize technology, understand data and write and communicate effectively. We need to be teaching all of these things. A national test on basics only may limit teaching and learning in the process skill areas.

- Any effort toward a national assessment must consider the extent to which states are now assessing. We are concerned about the huge demand on instructional time that the many layers of tests and the frequency of current assessments now require. In Michigan, we have yearly statewide every-pupil testing in Reading, Math and Science at several grade intervals. Many districts employ competency testing for graduation, teacher-developed tests are used routinely, as well as the various national tests currently in place (NAEP, ACT, SAT).

- Comparisons of test data between school districts, states, and even other countries are often unproductive because they do not take regional and other differences into account. For example, urban (economically and educationally disadvantaged, ethnic differences, fewer educational resources) and suburban (middle class, professional, greater educational resources) school district comparisons demonstrate the need to develop guidelines for interpreting and using test results beyond raw or aggregate score comparisons of "how well" a district is doing. We subscribe to the philosophy that all students can and must learn essential skills, concepts and processes. Test results should be used to examine the extent to which educational equity and quality of learning has been enhanced, over time within a particular district or state, not in comparison to others who may have started at a different point on the scale due to certain advantages or disadvantages they have.

- Development of a national curriculum, or a set of national goals and objectives for education, must involve collaboration with not only the leading subject area specialists and general curriculum development experts, but also with the local school district teachers and administrators who are most closely associated with the success of our educational system.

We appreciate this opportunity to communicate our views on the national testing debate. Please do not hesitate to call on Middle Cities for information and assistance as your Committee debates this important issue.

Sincerely,

Michael A. Boulus
Executive Director

Marcia Leone
Legislative Information Director

MAB/ML/gu

Houghton Mifflin Company

One Beacon Street Boston Massachusetts 02108
(617) 725-5000 Cable Houghton

Michael E. Melody
Senior Vice President
College and Test
Publishing

STATEMENT OF MICHAEL E. MELODY
ON BEHALF OF HOUGHTON MIFFLIN COMPANY
SUBMITTED FOR INCLUSION IN THE RECORD
HEARING ON "NATIONAL TESTING: PROS AND CONS"
BEFORE THE SUBCOMMITTEE ON ELEMENTARY, SECONDARY AND VOCATIONAL EDUCATION
COMMITTEE ON EDUCATION AND LABOR
MARCH 14, 1991

Mr. Chairman and members of the Committee, my name is Michael E. Melody. I am Senior Vice President for College and Test Publishing for Houghton Mifflin Company. Houghton Mifflin Company is a publicly held, Massachusetts corporation which has been located in Boston for over 150 years. The company is a major publisher of textbooks and other educational materials, including assessment instruments, for schools and colleges, general interest and reference books for adults and young readers, and computer software for educational and business applications, for both domestic and international markets. I welcome this opportunity to contribute to your important work.

We employ over 1,400 people in Massachusetts and about that many more throughout the rest of the country. Houghton's annual Massachusetts payroll is more than $45 million, and we spend over $125 million locally with freelance writers, artists and photographers, printers, paper merchants, trucking companies, and other suppliers of goods and services each year.

Among our employees located elsewhere in the U.S. are nearly 400 people at two locations in Illinois -- our midwestern regional office in Geneva and The Riverside Publishing Company in Chicago.

The Riverside Publishing Company, a wholly owned subsidiary of Houghton Mifflin, publishes standardized tests in a variety of formats. Perhaps our best known program is the Iowa Test of Basic Skills which was first developed in 1935 under the direction of the staff of the College of Education at the University of Iowa in Iowa City.

Houghton Mifflin first published the Iowa tests in 1940, thereby making them available to schools nationwide. Our relationship with the University of Iowa, now through our subsidiary The Riverside Publishing Company, is one which has flourished for over fifty years. Millions of dollars and volumes of research data have been invested by both the University of Iowa and the publisher to developing and revising this testing program in order to keep it up to date in order to meet market requirements.

In addition, Riverside publishes the Integrated Literature and Language Arts Portfolio Program, a performance-based approach to evaluate student accomplishments in these critical areas of instruction and is developing a state program in performance-based testing of reading, language arts, and mathematics for Arizona. This project has attracted nationwide attention.

My primary purpose is to use this statement to provide a context for the consideration of the pros and cons of a national assessment program, from the experience base of a commercial publisher of standardized, multiple-choice tests. This is a particularly relevant perspective inasmuch as one rationale for a national assessment program is to compensate for the alleged deficiencies of such tests.

Houghton Mifflin and The Riverside Publishing Companies, and the other members of the AAP, remain committed to providing educational assessment instruments that are fair, valid, and reliable; to making use of technology; and to providing assessments with diverse formats, including multiple-choice, performance assessments and portfolio programs.

As Congress and others look at the issues involved with a new national assessment program, an understanding of current testing practices and the uses and limitations of testing should be helpful. Test publishers provide one -- but only one -- of the key elements in the instructional process. Our instruments provide information on individual performance, based on a sampling of skills and knowledge. Contrary to what is becoming the conventional wisdom, our tests can assess both basic skills and higher order skills, and can be administered in a variety of formats, such as multiple-choice, where the pupil selects an answer that is either right or wrong, or a "performance" format, where the student's performance is considered to be better or worse in terms of how that answer is arrived at and delivered.

The problems of education in America today can not be attributed simplistically to inadequate or inaccurate information derived from standardized tests. Standardized tests are not the only source of information on how well a student or a school system is doing. We have not made that claim. To do so would overstate the power of a test and undervalue the role of the teacher and the school system to provide relevant and accurate information.

Test-derived information can be used for many purposes. We believe that the most important use is in diagnosing individual strengths and weaknesses to improve the instruction of that pupil. Scores can also be used to determine how an individual compares to students nationwide as well as whether that individual has mastered local, state, or national instructional objectives. Individual scores can be used by the local and state education authorities for their own accountability purposes, to evaluate the results of the effectiveness of instruction, and to evaluate and improve curriculum.

To assure that tests are used effectively and appropriately, publishers provide specific guidelines and suggestions for the interpretation of test scores and how to present the scores in ways that both pupils and parents will find helpful; how to diagnose specific strengths and weaknesses; and how to use test results to improve classroom instruction. We provide in-service support to teachers and test administrators. These efforts are in accordance with the Standards for Educational and Psychological Testing and the Code of Fair Testing Practices in Education. (Copy attached)

To assure the school systems, parents, and the public of the quality of the tests being used, publishers also document the reliability and validity of their tests, and support their tests through extensive research and development efforts. Riverside's test authors are eminent educators and psychometricians with expertise not only in test construction, but also in curriculum development, instructional application, and the psychological implications of testing in school and university settings. Over fifty years of research at the University of Iowa and over 300 research studies have produced information about test construction that is the foundation of excellence on which Riverside's tests are built.

These same standards that have been applied to multiple-choice tests are also appropriate for performance assessments. It is important to stress that these formats are complementary. Performance assessments should not be considered a panacea, nor as a mutually-exclusive alternative to multiple-choice tests -- each serves special purposes and has unique advantages and disadvantages.

Performance assessments are not new. Teachers have traditionally evaluated writing skills by grading papers that students have written. Riverside's Iowa Tests of Basic Skills and Tests of Achievement and Proficiency include Writing supplements for grades 3-12 that have standardized this process by giving students a standard set of topics and standard guidelines for completing the essays. The essays are then scored using a standard procedure. Since all students follow similar directions and are graded in a consistent manner, the score for each writer's essay can be compared to national norms.

As mentioned earlier, Riverside also is developing a statewide performance assessment for Arizona. It will complement the Iowa Tests of Basic Skills, enabling assessment of progress over the school year from both a national and local perspective. The new test will be administered to grades three, eight, and twelve in the spring semester. It requires written responses to questions, including short answer and essay responses in the reading and writing sections, and graph and chart constructions in the math section. Arizona teachers will be trained to score the new tests, to assure that the results will be fair and comparable. Dr. Monty Neill of FairTest spoke positively of this project in his comments.

While it is true that performance assessments are not new, it is also important to point out that the use of this format for a wide range of subject areas in a high stakes context, where the results are to be used for comparisons of individual students and for system accountability, is breaking new ground. Accordingly, these expanded uses for performance assessments should be approached very cautiously to assure that they do meet the essential requirements of fairness, validity, and reliability.

Which format is used, of course, begs the question of what is being assessed. It is unfair to criticize multiple choice tests for not providing information on how a pupil is performing on national standards when those standards do not yet exist. In the absence of a national consensus on a standard for what should be taught and tested, standardized tests that compare a pupil to students nationwide have been the only objective and accurate appraisal of achievement beyond the local curriculum. We are now in a transition period, during which a consensus is growing that minimum national educational standards are appropriate and necessary. But that consensus is not yet attained nor has it been tested.

Is there a national curriculum now because publishers sell their texts and tests nationwide? We believe not. There still remains considerable diversity around the country on curriculum and the scope and sequence of instruction. Recent advances in electronic publishing also make it far easier and more cost-effective for us to accommodate the demands of our customers to respond to their needs. It also makes it easier for us to respond to rapidly changing national educational movements and the resultant market demands.

To provide one example, in 1988, the Bradley Commission on History in Schools released its report, "Building a History Curriculum: Guidelines for Teaching History in Schools," recommendations intended to improve the teaching of history as the core of the social studies. Houghton Mifflin Company responded to the Bradley Commission's guidelines with Houghton Mifflin Social Studies, 1991, a kindergarten through eighth-grade textbook series. Through this program, Houghton Mifflin met the Bradley Commission's guidelines and put its theories into practice. The books have already been adopted by California, Arkansas, Oregon, Indiana, and West Virginia, but it would be premature to say that it has nationwide acceptance. If there were to be a national assessment on social studies, would it reflect the Bradley Commission's guidelines or traditional social studies instruction? Would two assessments be needed, so that systems that have adopted the guidelines will not be disadvantaged? Would assessment be delayed until there was a national consensus on the Bradley guidelines? Should there be a national consensus on the adoption of the Bradley guidelines? These are very practical considerations that must be part of the decision-making process toward a national examination system.

As we know, similar situations exist in mathematics and science as the result of national standards being developed by the National Council Teachers of Mathematics, the Mathematics Science Education Board, and the American Association for the Advancement of Science, as pointed out by Mr. Shanker.

Our authors and editors do go through a consensus process in developing our tests. We consider:

- current emphases in instructional materials, such as textbooks,

- recommendations from national curriculum committees and teacher training specialists,

- critical evaluations and suggestions of classroom teachers and school administrators who use the tests,

- social utility studies in relevant curricular areas,

- comprehensive item tryouts and research studies to determine frequency of error, particularly in language and mathematics,

- independent reviews by professionals from a variety of cultural groups to assess the fairness and appropriateness of items relative to demographics, race, and sex; and

° topic importance, biased on authoritative judgment, instructional trends, and public opinion.

Other test publishers go through comparable processes. The diversity and dynamic process of publishing instructional materials also mitigate against homogeneity and minimize the chances of a close alignment of texts and nationally-available test. Factors against such alignment include that fact that there are many more textbook publishers than there are test publishers, texts and tests are being constantly revised in differing cycles, and tests cover multiple subject areas for which there are likely going to be multiple texts from different publishers. Tests, however, may be selected to more closely correlate to the local curriculum, which may facilitate a school system s use of the test to evaluate progress on local as well as national objectives.

Finally, let us recognize that over the past twenty years there have been real gains in basic reading and numeracy skills, with the greatest gains being realized by minority students. Minimum competency is now nearly universal among high school graduates. That does not mean complacency about higher order skills or the condition of drop-outs, but there is cause to accept that there have been successes. These gains have been recorded on numerous tests, not just on standardized, multiple-choice tests. If minimum national standards are established, and if those standards position our students to meet the challenges of an international economy, perhaps in a few years our assessments will show comparable gains to those shown on basic skills.

To assist the Committee s deliberations, I am submitting for inclusion with my statement materials which The Riverside Publishing Company has developed regarding performance and multiple-choice testing formats. Although I can speak only for Houghton Mifflin Company and The Riverside Publishing Company, I am confident that I speak for the commercial textbook and test publishers in general when I say that the industry stands ready to share our publishing expertise with the committee as you strive for a solution to the critical issues we are all addressing.

1909h

CODE OF FAIR TESTING PRACTICES IN EDUCATION

Prepared by the Joint Committee on Testing Practices

The Code of Fair Testing Practices in Education states the major obligations to test takers of professionals who develop or use educational tests. The Code is meant to apply broadly to the use of tests in education (admissions, educational assessment, educational diagnosis, and student placement). The Code is not designed to cover employment testing, licensure or certification testing, or other types of testing. Although the Code has relevance to many types of educational tests, it is directed primarily at professionally developed tests such as those sold by commercial test publishers or used in formally administered testing programs. The Code is not intended to cover tests made by individual teachers for use in their own classrooms.

The Code addresses the roles of test developers and test users separately. Test users are people who select tests, commission test development services, or make decisions on the basis of test scores. Test developers are people who actually construct tests as well as those who set policies for particular testing programs. The roles may, of course, overlap as when a state education agency commissions test development services, sets policies that control the test development process, and makes decisions on the basis of the test scores.

• • • • • • • • • • •

The Code has been developed by the Joint Committee on Testing Practices, a cooperative effort of several professional organizations, that has as its aim the advancement, in the public interest, of the quality of testing practices. The Joint Committee was initiated by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. In addition to these three groups, the American Association for Counseling and Development/Association for Measurement and Evaluation in Counseling and Development, and the American Speech-Language-Hearing Association are now also sponsors of the Joint Committee.

This is not copyrighted material. Reproduction and dissemination are encouraged. Please cite this document as follows:

Code of Fair Testing Practices in Education. (1988) Washington, D.C.: Joint Committee on Testing Practices. (Mailing Address: Joint Committee on Testing Practices, American Psychological Association, 1200 17th Street, NW, Washington, D.C. 20036.)

Code of Fair Testing Practices in Education

The Code presents standards for educational test developers and users in four areas:

- **A.** Developing/Selecting Tests
- **B.** Interpreting Scores
- **C.** Striving for Fairness
- **D.** Informing Test Takers

Organizations, institutions, and individual professionals who endorse the Code commit themselves to safeguarding the rights of test takers by following the principles listed. The Code is intended to be consistent with the relevant parts of the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985). However, the Code differs from the Standards in both audience and purpose. The Code is meant to be understood by the general public; it is limited to educational tests; and the primary focus is on those issues that affect the proper use of tests. The Code is not meant to add new principles over and above those in the Standards or to change the meaning of the Standards. The goal is rather to represent the spirit of a selected portion of the Standards in a way that is meaningful to test takers and/or their parents or guardians. It is the hope of the Joint Committee that the Code will also be judged to be consistent with existing codes of conduct and standards of other professional groups who use educational tests.

A Developing/Selecting Appropriate Tests*

Test developers should provide the information that test users need to select appropriate tests.	**Test users should select tests that meet the purpose for which they are to be used and that are appropriate for the intended test-taking populations.**
Test Developers Should:	**Test Users Should:**
1. Define what each test measures and what the test should be used for. Describe the population(s) for which the test is appropriate.	1. First define the purpose for testing and the population to be tested. Then, select a test for that purpose and that population based on a thorough review of the available information.
2. Accurately represent the characteristics, usefulness, and limitations of tests for their intended purposes.	2. Investigate potentially useful sources of information, in addition to test scores, to corroborate the information provided by tests.
3. Explain relevant measurement concepts as necessary for clarity at the level of detail that is appropriate for the intended audience(s).	3. Read the materials provided by test developers and avoid using tests for which unclear or incomplete information is provided.
4. Describe the process of test development. Explain how the content and skills to be tested were selected.	4. Become familiar with how and when the test was developed and tried out.
5. Provide evidence that the test meets its intended purpose(s).	5. Read independent evaluations of a test and of possible alternative measures. Look for evidence required to support the claims of test developers.
6. Provide either representative samples or complete copies of test questions, directions, answer sheets, manuals, and score reports to qualified users.	6. Examine specimen sets, disclosed tests or samples of questions, directions, answer sheets, manuals, and score reports before selecting a test.
7. Indicate the nature of the evidence obtained concerning the appropriateness of each test for groups of different racial, ethnic, or linguistic backgrounds who are likely to be tested.	7. Ascertain whether the test content and norms group(s) or comparison group(s) are appropriate for the intended test takers.
8. Identify and publish any specialized skills needed to administer each test and to interpret scores correctly.	8. Select and use only those tests for which the skills needed to administer the test and interpret scores correctly are available.

---

*Many of the statements in the Code refer to the selection of existing tests. However, in customized testing programs test developers are engaged to construct new tests. In those situations, the test development process should be designed to help ensure that the completed tests will be in compliance with the Code.

B Interpreting Scores

Test developers should help users interpret scores correctly.

Test Developers Should:

9. Provide timely and easily understood score reports that describe test performance clearly and accurately. Also explain the meaning and limitations of reported scores.
10. Describe the population(s) represented by any norms or comparison group(s), the dates the data were gathered, and the process used to select the samples of test takers.
11. Warn users to avoid specific, reasonably anticipated misuses of test scores.
12. Provide information that will help users follow reasonable procedures for setting passing scores when it is appropriate to use such scores with the test.
13. Provide information that will help users gather evidence to show that the test is meeting its intended purpose(s).

Test users should interpret scores correctly.

Test Users Should:

9. Obtain information about the scale used for reporting scores, the characteristics of any norms or comparison group(s), and the limitations of the scores.
10. Interpret scores taking into account any major differences between the norms or comparison groups and the actual test takers. Also take into account any differences in test administration practices or familiarity with the specific questions in the test.
11. Avoid using tests for purposes not specifically recommended by the test developer unless evidence is obtained to support the intended use.
12. Explain how any passing scores were set and gather evidence to support the appropriateness of the scores.
13. Obtain evidence to help show that the test is meeting its intended purpose(s).

C Striving for Fairness

Test developers should strive to make tests that are as fair as possible for test takers of different races, gender, ethnic backgrounds, or handicapping conditions.

Test Developers Should:

14. Review and revise test questions and related materials to avoid potentially insensitive content or language.
15. Investigate the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Enact procedures that help to ensure that differences in performance are related primarily to the skills under assessment rather than to irrelevant factors.
16. When feasible, make appropriately modified forms of tests or administration procedures available for test takers with handicapping conditions. Warn test users of potential problems in using standard norms with modified tests or administration procedures that result in non-comparable scores.

Test users should select tests that have been developed in ways that attempt to make them as fair as possible for test takers of different races, gender, ethnic backgrounds, or handicapping conditions.

Test Users Should:

14. Evaluate the procedures used by test developers to avoid potentially insensitive content or language.
15. Review the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Evaluate the extent to which performance differences may have been caused by inappropriate characteristics of the test.
16. When necessary and feasible, use appropriately modified forms of tests or administration procedures for test takers with handicapping conditions. Interpret standard norms with care in the light of the modifications that were made.

D Informing Test Takers

Under some circumstances, test developers have direct communication with test takers. Under other circumstances, test users communicate directly with test takers. Whichever group communicates directly with test takers should provide the information described below.

Test Developers or Test Users Should:

17. When a test is optional, provide test takers or their parents/guardians with information to help them judge whether the test should be taken, or if an available alternative to the test should be used.
18. Provide test takers the information they need to be familiar with the coverage of the test, the types of question formats, the directions, and appropriate test-taking strategies. Strive to make such information equally available to all test takers.

Under some circumstances, test developers have direct control of tests and test scores. Under other circumstances, test users have such control. Whichever group has direct control of tests and test scores should take the steps described below.

Test Developers or Test Users Should:

19. Provide test takers or their parents/guardians with information about rights test takers may have to obtain copies of tests and completed answer sheets, retake tests, have tests rescored, or cancel scores.
20. Tell test takers or their parents/guardians how long scores will be kept on file and indicate to whom and under what circumstances test scores will or will not be released.
21. Describe the procedures that test takers or their parents/guardians may use to register complaints and have problems resolved.

* * *

Note: The membership of the Working Group that developed the Code of Fair Testing Practices in Education and of the Joint Committee on Testing Practices that guided the Working Group was as follows:

Theodore P. Bartell
John R. Bergan
Esther E. Diamond
Richard P. Duran
Lorraine D. Eyde
Raymond D. Fowler
John J. Fremer
(Co-chair, JCTP and Chair, Code Working Group)
Edmund W. Gordon
Jo-Ida C. Hansen
James B. Lingwall
George F. Madaus
(Co-chair, JCTP)
Kevin L. Moreland
Jo-Ellen V. Perez
Robert J. Solomon
John T. Stewart
Carol Kehr Tittle
(Co-chair, JCTP)
Nicholas A. Vacc
Michael J. Zieky
Debra Boltas and Wayne Camara of the American Psychological Association served as staff liaisons.

Additional copies of the Code may be obtained from the National Council on Measurement in Education, 1230 Seventeenth Street, NW, Washington, D.C. 20036. Single copies are free.